

Snedden, S.  
091881569

09/881569

L1 FILE 'REGISTRY' ENTERED AT 14:29:32 ON 15 NOV 2002  
O S BTS.PCGAVMILYFHKWR/SQSP

L2 O S PCGAVMILY/SQSP

FILE 'HOME' ENTERED AT 14:32:35 ON 15 NOV 2002



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 15, 2002, 11:43:10 : Search time 35 seconds  
(without alignments)  
68.529 Million cell updates/sec

Title: US-09-881-569A-2  
Perfect score: 102  
Sequence: 1 BTRSPCGAVMILYFKMR 18

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Minimum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : A\_Geneseq\_101002:\*
- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*
  - 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*
  - 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*
  - 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*
  - 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*
  - 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*
  - 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*
  - 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*
  - 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*
  - 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*
  - 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*
  - 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*
  - 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*
  - 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*
  - 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*
  - 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*
  - 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*
  - 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*
  - 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*
  - 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*
  - 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*
  - 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*
  - 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID       | Description          |
|------------|-------|-------------|--------|----|----------|----------------------|
| 1          | 50    | 49.0        | 281    | 21 | AAG06878 | Arabidopsis thaliana |
| 2          | 50    | 49.0        | 296    | 21 | AAG06877 | Arabidopsis thaliana |
| 3          | 46    | 45.1        | 99     | 21 | AAG26837 | Zea mays protein f   |
| 4          | 44    | 43.1        | 53     | 23 | ABP04271 | Human ORFX protein   |
| 5          | 44    | 43.1        | 893    | 23 | AA014530 | Protein of Me2 (M    |
| 6          | 43    | 42.2        | 954    | 23 | AAU78464 | Sphingolipid ceram   |
| 7          | 43    | 42.2        | 992    | 23 | AAU78470 | Sphingolipid ceram   |
| 8          | 42    | 41.2        | 57     | 21 | AAB14174 | Tissue factor path   |
| 9          | 42    | 41.2        | 223    | 20 | AAY35526 | Amino acid sequenc   |
| 10         | 41    | 40.2        | 58     | 6  | AAP50700 | Sequence of Gly(15   |

|    |      |      |      |    |           |                      |
|----|------|------|------|----|-----------|----------------------|
| 11 | 41   | 40.2 | 504  | 19 | AAW35008  | Vibrio harveyi end   |
| 12 | 41   | 40.2 | 544  | 19 | AAW59445  | Hordeum vulgare ML   |
| 13 | 41   | 40.2 | 562  | 22 | AAU42094  | Protonibacterium     |
| 14 | 41   | 40.2 | 575  | 23 | ABB48756  | Listeria monocytog   |
| 15 | 40   | 39.2 | 60   | 20 | AAV08613  | Inter-alpha-trypsi   |
| 16 | 40   | 39.2 | 74   | 21 | AAAG45083 | Arabidopsis thaliana |
| 17 | 40   | 39.2 | 75   | 22 | AAAG7642  | Human immune/haema   |
| 18 | 40   | 39.2 | 76   | 21 | AAAG24554 | Arabidopsis thaliana |
| 19 | 40   | 39.2 | 80   | 21 | AAAG45059 | Arabidopsis thaliana |
| 20 | 40   | 39.2 | 100  | 21 | AAAG12492 | Zea mays protein f   |
| 21 | 40   | 39.2 | 101  | 21 | AAAG23754 | Arabidopsis thaliana |
| 22 | 40   | 39.2 | 101  | 21 | AAAG33794 | Arabidopsis thaliana |
| 23 | 40   | 39.2 | 118  | 21 | AAAG18532 | Zea mays protein f   |
| 24 | 40   | 39.2 | 138  | 21 | AAAG45082 | Arabidopsis thaliana |
| 25 | 40   | 39.2 | 140  | 21 | AAAG24553 | Arabidopsis thaliana |
| 26 | 40   | 39.2 | 144  | 21 | AAAG45057 | Arabidopsis thaliana |
| 27 | 40   | 39.2 | 187  | 21 | AAAG45053 | Arabidopsis thaliana |
| 28 | 40   | 39.2 | 253  | 21 | AAAG45052 | Arabidopsis thaliana |
| 29 | 40   | 39.2 | 476  | 22 | AAAB61988 | Wild-type Tn5 tran   |
| 30 | 40   | 39.2 | 629  | 22 | ABBS58883 | Drosophila melanog   |
| 31 | 39.5 | 38.7 | 736  | 20 | AAV07046  | Breast cancer asso   |
| 32 | 39.5 | 38.7 | 1157 | 23 | ABBA48681 | Listeria monocytog   |
| 33 | 39   | 38.2 | 39   | 23 | ABGA0377  | Human peptidase enco |
| 34 | 39   | 38.2 | 75   | 23 | ABP35281  | Human kinase-like    |
| 35 | 39   | 38.2 | 111  | 13 | AAAR22980 | Aprotinin analogue   |
| 36 | 39   | 38.2 | 113  | 21 | AAAB41073 | Human ORFX ORF837    |
| 37 | 39   | 38.2 | 113  | 23 | ABP33735  | Human ORF2708 prot   |
| 38 | 39   | 38.2 | 138  | 23 | ABP30313  | Streptococcus poly   |
| 39 | 39   | 38.2 | 143  | 23 | ABP27753  | Streptococcus poly   |
| 40 | 39   | 38.2 | 217  | 23 | ABP26338  | Streptococcus poly   |
| 41 | 39   | 38.2 | 270  | 22 | ABG05544  | Novel human diagno   |
| 42 | 39   | 38.2 | 282  | 22 | AAAG84924 | Shrimp white spot    |
| 43 | 39   | 38.2 | 284  | 23 | ABBS5547  | Lactococcus lactis   |
| 44 | 39   | 38.2 | 697  | 22 | ABG16522  | Novel human diagno   |
| 45 | 39   | 38.2 | 1162 | 22 | ABG05052  | Novel human diagno   |

ALIGNMENTS

|   |                                   |
|---|-----------------------------------|
| RESULT 1  |                                   |
| AAG06878  |                                   |
| ID AAG06878   | standard; Protein; 281 AA.        |
| AC AAG06878;  |                                   |
| XX  |                                   |
| DF 17-OCT-2000  | (first entry)                     |
| XX  |                                   |
| DE Arabidopsis thaliana   | protein fragment SEQ ID NO: 3815. |
| XX  |                                   |
| KW Protein identification; signal transduction pathway; metabolic pathway;  |                                   |
| KW hybridisation assay; genetic mapping; gene expression control; promoter; |                                   |
| KW termination sequence.  |                                   |
| XX  |                                   |
| OS Arabidopsis thaliana.  |                                   |
| XX  |                                   |
| PN EP1033405-A2.  |                                   |
| XX  |                                   |
| PD 06-SEP-2000.   |                                   |
| XX  |                                   |
| PF 25-FEB-2000; 2000EP-0301439.   |                                   |
| XX  |                                   |
| XX 25-FEB-1999;   | 99US-0121825.                     |
| PR 05-MAR-1999;   | 99US-0123180.                     |
| PR 09-MAR-1999;   | 99US-0123548.                     |
| PR 23-MAR-1999;   | 99US-0125788.                     |
| PR 25-MAR-1999;   | 99US-0126264.                     |
| PR 29-MAR-1999;   | 99US-0126785.                     |
| PR 01-APR-1999;   | 99US-0127462.                     |
| PR 06-APR-1999;   | 99US-0128234.                     |
| PR 08-APR-1999;   | 99US-0128714.                     |
| PR 16-APR-1999;   | 99US-0129845.                     |
| PR 19-APR-1999;   | 99US-0130077.                     |

PR 21-APR-1999; 99US-0130449.  
 PR 23-APR-1999; 99US-0130510.  
 PR 23-APR-1999; 99US-0130891.  
 PR 28-APR-1999; 99US-0131449.  
 PR 30-APR-1999; 99US-0132048.  
 PR 30-APR-1999; 99US-0132407.  
 PR 04-MAY-1999; 99US-0132484.  
 PR 05-MAY-1999; 99US-0132485.  
 PR 06-MAY-1999; 99US-0132486.  
 PR 06-MAY-1999; 99US-0132487.  
 PR 07-MAY-1999; 99US-0132863.  
 PR 11-MAY-1999; 99US-0134256.  
 PR 14-MAY-1999; 99US-0134218.  
 PR 14-MAY-1999; 99US-0134219.  
 PR 14-MAY-1999; 99US-0134221.  
 PR 14-MAY-1999; 99US-0134370.  
 PR 18-MAY-1999; 99US-0134768.  
 PR 19-MAY-1999; 99US-0134941.  
 PR 20-MAY-1999; 99US-0135124.  
 PR 21-MAY-1999; 99US-0135353.  
 PR 24-MAY-1999; 99US-0135629.  
 PR 25-MAY-1999; 99US-0136021.  
 PR 27-MAY-1999; 99US-0136392.  
 PR 28-MAY-1999; 99US-0136782.  
 PR 01-JUN-1999; 99US-0137222.  
 PR 03-JUN-1999; 99US-0137528.  
 PR 04-JUN-1999; 99US-0137502.  
 PR 07-JUN-1999; 99US-0137724.  
 PR 08-JUN-1999; 99US-0138094.  
 PR 10-JUN-1999; 99US-0138540.  
 PR 10-JUN-1999; 99US-0138847.  
 PR 14-JUN-1999; 99US-0139119.  
 PR 16-JUN-1999; 99US-0139452.  
 PR 16-JUN-1999; 99US-0139453.  
 PR 17-JUN-1999; 99US-0139492.  
 PR 18-JUN-1999; 99US-0139454.  
 PR 18-JUN-1999; 99US-0139455.  
 PR 18-JUN-1999; 99US-0139456.  
 PR 18-JUN-1999; 99US-0139457.  
 PR 18-JUN-1999; 99US-0139458.  
 PR 18-JUN-1999; 99US-0139459.  
 PR 18-JUN-1999; 99US-0139460.  
 PR 18-JUN-1999; 99US-0139461.  
 PR 18-JUN-1999; 99US-0139462.  
 PR 18-JUN-1999; 99US-0139463.  
 PR 18-JUN-1999; 99US-0139750.  
 PR 18-JUN-1999; 99US-0139763.  
 PR 21-JUN-1999; 99US-0139817.  
 PR 22-JUN-1999; 99US-0139899.  
 PR 23-JUN-1999; 99US-0140353.  
 PR 23-JUN-1999; 99US-0140354.  
 PR 24-JUN-1999; 99US-0140695.  
 PR 28-JUN-1999; 99US-0140823.  
 PR 29-JUN-1999; 99US-0140991.  
 PR 30-JUN-1999; 99US-0141287.  
 PR 01-JUL-1999; 99US-0141842.  
 PR 01-JUL-1999; 99US-0142154.  
 PR 02-JUL-1999; 99US-0142055.  
 PR 06-JUL-1999; 99US-0142390.  
 PR 08-JUL-1999; 99US-0142803.  
 PR 09-JUL-1999; 99US-0142920.  
 PR 12-JUL-1999; 99US-0142977.  
 PR 13-JUL-1999; 99US-0143542.  
 PR 14-JUL-1999; 99US-0143624.  
 PR 15-JUL-1999; 99US-0144005.  
 PR 16-JUL-1999; 99US-0144085.  
 PR 16-JUL-1999; 99US-0144086.  
 PR 19-JUL-1999; 99US-0144325.  
 PR 19-JUL-1999; 99US-0144331.  
 PR 19-JUL-1999; 99US-0144332.  
 PR 19-JUL-1999; 99US-0144333.  
 PR 19-JUL-1999; 99US-0144334.  
 PR 19-JUL-1999; 99US-0144335.

PR 20-JUL-1999; 99US-0144352.  
 PR 20-JUL-1999; 99US-0144632.  
 PR 20-JUL-1999; 99US-0144884.  
 PR 21-JUL-1999; 99US-0144814.  
 PR 21-JUL-1999; 99US-0145086.  
 PR 21-JUL-1999; 99US-0145088.  
 PR 22-JUL-1999; 99US-0145085.  
 PR 22-JUL-1999; 99US-0145087.  
 PR 22-JUL-1999; 99US-0145089.  
 PR 22-JUL-1999; 99US-0145192.  
 PR 23-JUL-1999; 99US-0145145.  
 PR 23-JUL-1999; 99US-0145218.  
 PR 23-JUL-1999; 99US-0145224.  
 PR 26-JUL-1999; 99US-0145276.  
 PR 27-JUL-1999; 99US-0145913.  
 PR 27-JUL-1999; 99US-0145918.  
 PR 27-JUL-1999; 99US-0145919.  
 PR 28-JUL-1999; 99US-0145951.  
 PR 02-AUG-1999; 99US-0146386.  
 PR 02-AUG-1999; 99US-0146388.  
 PR 02-AUG-1999; 99US-0146389.  
 PR 03-AUG-1999; 99US-0147038.  
 PR 04-AUG-1999; 99US-0147204.  
 PR 04-AUG-1999; 99US-0147302.  
 PR 05-AUG-1999; 99US-0147192.  
 PR 05-AUG-1999; 99US-0147260.  
 PR 06-AUG-1999; 99US-0147303.  
 PR 06-AUG-1999; 99US-0147416.  
 PR 09-AUG-1999; 99US-0147493.  
 PR 10-AUG-1999; 99US-0147935.  
 PR 11-AUG-1999; 99US-0148171.  
 PR 12-AUG-1999; 99US-0148319.  
 PR 13-AUG-1999; 99US-0148565.  
 PR 13-AUG-1999; 99US-0148684.  
 PR 16-AUG-1999; 99US-0149368.  
 PR 17-AUG-1999; 99US-0149175.  
 PR 18-AUG-1999; 99US-0149426.  
 PR 20-AUG-1999; 99US-0149722.  
 PR 20-AUG-1999; 99US-0149723.  
 PR 20-AUG-1999; 99US-0149929.  
 PR 23-AUG-1999; 99US-0149902.  
 PR 23-AUG-1999; 99US-0149930.  
 PR 25-AUG-1999; 99US-0150566.  
 PR 26-AUG-1999; 99US-0150884.  
 PR 27-AUG-1999; 99US-0151065.  
 PR 27-AUG-1999; 99US-0151066.  
 PR 27-AUG-1999; 99US-0151080.  
 PR 30-AUG-1999; 99US-0151303.  
 PR 31-AUG-1999; 99US-0151438.  
 PR 01-SEP-1999; 99US-0151930.  
 PR 07-SEP-1999; 99US-0152363.  
 PR 10-SEP-1999; 99US-0153070.  
 PR 13-SEP-1999; 99US-0153758.  
 PR 15-SEP-1999; 99US-0154018.  
 PR 16-SEP-1999; 99US-0154039.  
 PR 20-SEP-1999; 99US-0154779.  
 PR 22-SEP-1999; 99US-0155139.  
 PR 23-SEP-1999; 99US-0155486.  
 PR 24-SEP-1999; 99US-0155659.  
 PR 28-SEP-1999; 99US-0156458.  
 PR 29-SEP-1999; 99US-0156596.  
 PR 04-OCT-1999; 99US-0157117.  
 PR 05-OCT-1999; 99US-0157753.  
 PR 06-OCT-1999; 99US-0157865.  
 PR 07-OCT-1999; 99US-0158029.  
 PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0159293.  
 PR 13-OCT-1999; 99US-0159294.  
 PR 13-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.



PR 22-JUL-1999; 99US-0145087.  
 PR 22-JUL-1999; 99US-0145089.  
 PR 22-JUL-1999; 99US-0145192.  
 PR 23-JUL-1999; 99US-0145145.  
 PR 23-JUL-1999; 99US-0145218.  
 PR 23-JUL-1999; 99US-0145224.  
 PR 26-JUL-1999; 99US-0145276.  
 PR 27-JUL-1999; 99US-0145913.  
 PR 27-JUL-1999; 99US-0145918.  
 PR 27-JUL-1999; 99US-0145919.  
 PR 28-JUL-1999; 99US-0145951.  
 PR 02-AUG-1999; 99US-0146386.  
 PR 02-AUG-1999; 99US-0146388.  
 PR 02-AUG-1999; 99US-0146389.  
 PR 03-AUG-1999; 99US-0147038.  
 PR 04-AUG-1999; 99US-0147204.  
 PR 04-AUG-1999; 99US-0147302.  
 PR 05-AUG-1999; 99US-0147192.  
 PR 05-AUG-1999; 99US-0147260.  
 PR 06-AUG-1999; 99US-0147303.  
 PR 06-AUG-1999; 99US-0147416.  
 PR 09-AUG-1999; 99US-0147493.  
 PR 09-AUG-1999; 99US-0147935.  
 PR 10-AUG-1999; 99US-0148171.  
 PR 11-AUG-1999; 99US-0148319.  
 PR 12-AUG-1999; 99US-0148341.  
 PR 13-AUG-1999; 99US-0148565.  
 PR 13-AUG-1999; 99US-0148684.  
 PR 16-AUG-1999; 99US-0149368.  
 PR 17-AUG-1999; 99US-0149175.  
 PR 18-AUG-1999; 99US-0149426.  
 PR 20-AUG-1999; 99US-0149722.  
 PR 20-AUG-1999; 99US-0149723.  
 PR 20-AUG-1999; 99US-0149929.  
 PR 23-AUG-1999; 99US-0149902.  
 PR 23-AUG-1999; 99US-0149930.  
 PR 25-AUG-1999; 99US-0150566.  
 PR 26-AUG-1999; 99US-0150884.  
 PR 27-AUG-1999; 99US-0151065.  
 PR 27-AUG-1999; 99US-0151066.  
 PR 27-AUG-1999; 99US-0151080.  
 PR 30-AUG-1999; 99US-0151303.  
 PR 31-AUG-1999; 99US-0151438.  
 PR 01-SEP-1999; 99US-0151930.  
 PR 07-SEP-1999; 99US-0152363.  
 PR 10-SEP-1999; 99US-0153070.  
 PR 13-SEP-1999; 99US-0153758.  
 PR 15-SEP-1999; 99US-0154018.  
 PR 16-SEP-1999; 99US-0154039.  
 PR 20-SEP-1999; 99US-0154779.  
 PR 22-SEP-1999; 99US-0155139.  
 PR 23-SEP-1999; 99US-0155486.  
 PR 24-SEP-1999; 99US-0155659.  
 PR 28-SEP-1999; 99US-0156458.  
 PR 29-SEP-1999; 99US-0156596.  
 PR 04-OCT-1999; 99US-0157117.  
 PR 05-OCT-1999; 99US-0157753.  
 PR 06-OCT-1999; 99US-0157865.  
 PR 07-OCT-1999; 99US-0158029.  
 PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0159293.  
 PR 13-OCT-1999; 99US-0159294.  
 PR 13-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 14-OCT-1999; 99US-0159638.  
 PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.

PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 49.0%; Score 50; DB 21; Length 296;  
 Best Local Similarity 34.8%; Pred. No. 2.9;  
 Matches 8; Conservative 6; Mismatches 3; Indels 6; Gaps 1;

QY 2 TSZPCGAVMIL-----YFHWR 18  
 ||: ||::: | |:  
 Db 267 TSKGCGCLLVDEHFDRIYHWR 289

RESULT 3  
 AAG26837  
 ID AAG26837 standard; Protein; 99 AA.  
 AC AAG26837;  
 XX 17-OCT-2000 (first entry)  
 DT Zea mays protein fragment SEQ ID NO: 31441.  
 DE  
 XX Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 KM termination sequence; corn.  
 XX  
 OS Zea mays subsp. mays.  
 XX  
 PN EP1033405-A2.  
 XX  
 PD 06-SEP-2000.  
 XX  
 PF 25-FEB-2000; 2000EP-0301439.  
 XX  
 PR 25-FEB-1999; 99US-0121825.  
 PR 05-MAR-1999; 99US-0123180.  
 PR 09-MAR-1999; 99US-0123548.  
 PR 23-MAR-1999; 99US-0125788.  
 PR 25-MAR-1999; 99US-0126264.  
 PR 29-MAR-1999; 99US-0126785.  
 PR 01-APR-1999; 99US-0127462.  
 PR 06-APR-1999; 99US-0128234.  
 PR 08-APR-1999; 99US-0128714.  
 PR 16-APR-1999; 99US-0129845.  
 PR 19-APR-1999; 99US-0130077.  
 PR 21-APR-1999; 99US-0130449.  
 PR 23-APR-1999; 99US-0130510.  
 PR 23-APR-1999; 99US-0130891.  
 PR 28-APR-1999; 99US-0131449.  
 PR 30-APR-1999; 99US-0132048.  
 PR 30-APR-1999; 99US-0132407.  
 PR 04-MAY-1999; 99US-0132484.  
 PR 05-MAY-1999; 99US-0132485.  
 PR 06-MAY-1999; 99US-0132486.  
 PR 06-MAY-1999; 99US-0132487.  
 PR 07-MAY-1999; 99US-0132863.  
 PR 11-MAY-1999; 99US-0134256.  
 PR 14-MAY-1999; 99US-0134218.  
 PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.

PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.





XX Novel polycomb group nucleic acids and polypeptides from maize useful  
PT for producing transgenic plants and to repress expression or prevent  
PT the expression of a desired target gene in specific tissue in plants in  
PT vivo -  
PS Claim 8; Fig 2B; 80pp; English.  
XX  
CC The invention relates to plant genetic engineering. More specifically,  
CC the invention relates to isolated and purified polycomb group (PCG) genes  
CC from Zea mays (maize), termed Mez1 and Mez2 (Maize E(z)-like 1 and 2)  
CC genes comprising a sequence of 3180 and 3030 bp respectively, defined in  
CC the specification, or its conservatively modified and polymorphic  
CC variants. The PCG genes of the invention are useful in constructing  
CC recombinant expression cassettes which can be introduced into a desired  
CC plant, to produce a transgenic plant. Promoters derived from the Mez1 or  
CC Mez2 genes are useful to direct expression of heterologous sequences. The  
CC modified variants of the PCG genes are useful for generating or selecting  
CC antibodies immunoreactive to the non-variant polypeptide and to follow  
CC the segregation of chromosome regions and in marker assisted selection  
CC methods for crop improvement. Mez1 and Mez2 proteins can be used to  
CC repress the expression or prevent the expression of a desired target gene  
CC in specific tissue in a plant in vivo, to regulate homeotic gene  
CC expression in plants to create novel plants having improved agronomic  
CC traits. Protein derived from a PCG gene of the invention is useful as an  
CC immunogen for producing monoclonal or polyclonal antibodies which are  
CC useful for screening plants for the expression of a Mez1 or Mez2  
CC polypeptide, and for affinity chromatography for isolating a Mez1 or Mez2  
CC polypeptide. This sequence represents the protein of Mez 2 of the  
CC invention.  
XX  
SQ Sequence 893 AA;  
Query Match 43.1%; Score 44; DB 23; Length 893;  
Best Local Similarity 38.9%; Pred. No. 95;  
Matches 7; Conservative 4; Mismatches 7; Indels 0; Gaps 0;  
QY 1 BRSZPCGAVMILYFHKWR 18  
: :||| : | | :||  
Db 317 ENKKPCGHLCLRFPQWR 334  
RESULT 6  
AAU78464  
ID AAU78464 standard; Protein; 954 AA.  
XX  
AC AAU78464;  
XX  
DT 02-JUL-2002 (first entry)  
XX  
DE Sphingolipid ceramide deacylase.  
XX  
XX Sphingolipid ceramide deacylase; nootropic; neuroprotective; cytostatic;  
XX Shewanella alga.  
OS  
XX  
XX WO200226963-A1.  
PN  
XX  
PD 04-APR-2002.  
XX  
PF 26-SEP-2001; 2001WO-JP08344.  
XX  
PR 26-SEP-2000; 2000JP-0293181.  
XX  
PA (TAKI ) TAKARA SHUZO CO LTD.  
XX  
PI Ito M, Furusato M, Sueyoshi N;  
XX  
XX WPI; 2002-340011/37.  
DR  
DR N-PSDB; ABK47644.  
XX  
PT New polypeptide with sphingolipid ceramide deacylase activity and

PT encoded gene, applicable in studying structure and function of  
PT sphingolipid and its engineering for treatment of e.g.  
PT neurodegenerative diseases and leukaemia -  
XX  
PS Example 6; Page 59-64; 89pp; Japanese.  
XX  
CC The invention relates to a polypeptide with sphingolipid ceramide  
CC deacylase activity and its encoding nucleic acid. The polypeptide and its  
CC encoded gene are applicable in studying structure and function of  
CC sphingolipid and its engineering for treatment of e.g. neurodegenerative  
CC diseases, leukaemia and wounds. The present sequence represents the  
CC amino acid sequence of sphingolipid ceramide deacylase.  
XX  
SQ Sequence 954 AA;  
Query Match 42.2%; Score 43; DB 23; Length 954;  
Best Local Similarity 34.8%; Pred. No. 1.5e+02;  
Matches 8; Conservative 4; Mismatches 5; Indels 6; Gaps 1;  
QY 2 TSZPCGAVMILYFH-----KWR 18  
: : || :||| | || :||  
Db 929 SGEQAGAAALILYHHGDSNQKWK 951  
RESULT 7  
AAU78470  
ID AAU78470 standard; Protein; 992 AA.  
XX  
AC AAU78470;  
XX  
DT 02-JUL-2002 (first entry)  
XX  
DE Sphingolipid ceramide deacylase #2.  
XX  
KW Sphingolipid ceramide deacylase; nootropic; neuroprotective; cytostatic;  
KW vulnerable; neurodegenerative disease; leukaemia; wound; enzyme.  
XX  
OS Shewanella alga.  
XX  
PN WO200226963-A1.  
XX  
PD 04-APR-2002.  
XX  
PF 26-SEP-2001; 2001WO-JP08344.  
XX  
PR 26-SEP-2000; 2000JP-0293181.  
XX  
PA (TAKI ) TAKARA SHUZO CO LTD.  
XX  
PI Ito M, Furusato M, Sueyoshi N;  
XX  
XX WPI; 2002-340011/37.  
DR  
DR N-PSDB; ABK47651.  
XX  
PT New polypeptide with sphingolipid ceramide deacylase activity and  
PT encoded gene, applicable in studying structure and function of  
PT sphingolipid and its engineering for treatment of e.g.  
PT neurodegenerative diseases and leukaemia -  
XX  
PS Example 2; Page 76-78; 89pp; Japanese.  
XX  
CC The invention relates to a polypeptide with sphingolipid ceramide  
CC deacylase activity and its encoding nucleic acid. The polypeptide and its  
CC encoded gene are applicable in studying structure and function of  
CC sphingolipid and its engineering for treatment of e.g. neurodegenerative  
CC diseases, leukaemia and wounds. The present sequence represents the  
CC amino acid sequence of sphingolipid ceramide deacylase #2.  
XX  
SQ Sequence 992 AA;  
Query Match 42.2%; Score 43; DB 23; Length 992;  
Best Local Similarity 34.8%; Pred. No. 1.6e+02;  
Matches 8; Conservative 4; Mismatches 5; Indels 6; Gaps 1;

OY 2 TSZPCGAVMILYFHKWR 18  
DB 967 SGEQAGALILYHHGDSNOKWK 989

RESULT 8

ID AAB14174 standard; protein; 57 AA.

AC AAB14174;

DT 02-FEB-2001 (first entry)

DE Tissue factor pathway inhibitor precursor 2 fragment # 1.

XX Mucociliary dysfunction; mucus; sputum;  
XX chronic obstructive lung disease; chronic bronchitis; CB; Bronchiectasis;  
XX BE; asthma; cystic fibrosis; CF; bacterial infection; placental bikunin;  
XX Kunitz-type serine protease inhibitor; chronic sinusitis; glue ear;  
XX tissue factor pathway inhibitor precursor.

OS Unidentified.

PN WO200037099-A2.

PD 29-JUN-2000.

PF 22-DEC-1999; 99WO-GB04381.

PR 22-DEC-1998; 98US-0218913.

PR 17-NOV-1999; 99US-0441966.

PA (FARB ) BAYER AG.

PI Hall R, Poll CF, Newton BB, Taylor WJA;

DR WPI; 2000-452127/39.

PT Stimulating mucociliary clearance rate of mucus and sputum in lung  
PT airways for treating lung diseases such as cystic fibrosis and  
PT bronchitis involves administering a Kunitz-type serine protease  
PT inhibitor

XX Disclosure; Page 22; 173pp; English.

CC Mucociliary dysfunction is the inability of ciliated epithelium to clear  
CC mucus and sputum in lung airways. Mucociliary dysfunction is a serious  
CC complication of chronic obstructive lung diseases such as Chronic  
CC Bronchitis (CB), Bronchiectasis (BE), asthma and Cystic Fibrosis (CF).  
CC In addition, patients suffering from mucociliary dysfunction are  
CC susceptible to secondary bacterial infections. The protein sequence of  
CC human placental bikunin has been isolated (see AAB14159). Placental  
CC bikunin is a Kunitz-type serine protease inhibitor protein, which can  
CC stimulate the rate of mucociliary clearance of mucus and sputum in lung  
CC airways. Therefore, placental bikunin may be used for treating lung  
CC diseases such as CF, CB, BE, and chronic sinusitis and glue ear which are  
CC caused by retention and accumulation of mucus. The present sequence is a  
CC fragment of tissue factor pathway inhibitor precursor 2. The present  
CC sequence is a known Kunitz-like domain and was used to identify the  
CC Kunitz-domains of placental bikunin.

SO Sequence 57 AA;

Query Match 41.2%; Score 42; DB 21; Length 57;

Best Local Similarity 35.7%; Pred. No. 11;

Matches 5; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

OY 5 PCGAVMILYFHKWR 18  
DB 12 PCRALLLRYYYR 25

RESULT 9

ID AAY35526 standard; protein; 223 AA.

AC AAY35526;

DT 13-SEP-1999 (first entry)

DE Amino acid sequence of a Chlamydia pneumoniae protein.

XX Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;  
XX sinusitis; purulent otitis media; erythema nodosum; pharyngitis;  
XX vaccine; neutralising epitope.

OS Chlamydia pneumoniae.

PN WO9927105-A2.

PD 03-JUN-1999.

PF 20-NOV-1998; 98WO-IB01890.

PR 04-NOV-1998; 98US-0107078.

PR 21-NOV-1997; 97FR-0014673.

PA (GEST ) GENSET.

PI Griffais R;

DR WPI; 1999-357842/30.

PT Genome sequence of Chlamydia pneumoniae

XX Page 1278-1279; Disclosure; 1912pp; English.

CC AAY34584-Y35879 represent the proteins encoded by all the open reading  
CC frames in the complete genome (see AAX91990) of Chlamydia pneumoniae.  
CC C. pneumoniae causes respiratory disease such as pneumonia and  
CC bronchitis and is thought to be a contributing factor in heart  
CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema  
CC nodosum or pharyngitis. The polypeptides encoded by the open reading  
CC frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in  
CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae  
CC nucleotide sequences can also be used as immunogenic compositions,  
CC especially where the vector directs the expression of a neutralising  
CC epitope of C. pneumoniae.

SO Sequence 223 AA;

Query Match 41.2%; Score 42; DB 20; Length 223;

Best Local Similarity 35.3%; Pred. No. 48;

Matches 6; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

OY 2 TSZPCGAVMILYFHKWR 18

DB 7 TTEPCSSIHISLNDWR 23

RESULT 10

ID AAP50700 standard; protein; 58 AA.

AC AAP50700;

DT 28-NOV-1991 (first entry)

DE Sequence of GLY(15)-Aprotinin.

XX Protease inhibitor; antitrypsin activity; shock therapy;  
XX kallikrein inhibitor; elastase inhibitor; arthritis therapy.

OS Sus scrofa domestica.

PN EPI32732-A.  
XX  
PD 13-FEB-1985.  
XX  
PF 14-JUL-1984; 84EP-0108293.  
XX  
PR 03-NOV-1983; 83DE-3339693.  
PR 28-JUL-1983; 83DE-3327277.  
XX  
PA (FARB ) BAYER AG.  
XX  
PI Tschesche H, Wenzel H, Schmuck R, Schnabel E;  
XX  
DR WPI; 1985-039382/07.  
XX  
PT New aprotinin derivs. with 15-lysine replaced by aminoacid - and  
PT new intermediate, useful as protease inhibitors, e.g. for  
PT treating shock  
XX  
PS Claim 1; Page 1; 51pp; German.  
XX  
CC The inventors claim homologues of aprotinin in which 15-Lys has been  
CC replaced by one of (all L forms where applicable) Gly,Ala,Val,Leu,  
CC Ile,Met,Arg, alpha-aminobutyric acid, norvaline, Nle, dehydroalanine  
CC or homoserine. Also new are the intermediate derivs. in which  
CC (1) 15-Lys is replaced by one of the specified AAs or their alkyl  
CC esters and any additional COOH in side chains or terminal posn. are  
CC partly or completely esterified; (2) the bond between 15-Lys and 16-  
CC Ala is open and only 15-Lys COOH is in free form, all other side  
CC chain COOH (3- and 50-Asp; 7- and 49-Glu) and C-terminal COOH (58-  
CC Ala) are present as alkyl esters; (3) the 15/16 bond is broken;  
CC 15-Lys is absent, the side chain and terminal COOH are esterified  
CC and only 14-Cys has a COOH in free form.  
XX  
SQ Sequence 58 AA;  
XX  
Query Match 40.2%; Score 41; DB 6; Length 58;  
Best Local Similarity 63.6%; Pred. No. 17;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
QY 5 PCGAVMILYFH 15  
| | | | : | | | |  
Db 13 PCGARIRIRFY 23  
XX  
RESULT 11  
AAW35008  
ID AAW35008 standard; Protein; 504 AA.  
XX  
AC AAW35008;  
XX  
XX 21-MAY-1998 (first entry)  
XX  
Vibrio harveyi endoglucanase.  
XX  
KM Endoglucanase; cellulase; carboxymethylcellulose; cellulose;  
KM biomass; beta-1,4-glycosidic bond; hydrolysis; saccharification;  
KM thermostable enzyme; thermophilic; glycosidase.  
XX  
OS Vibrio harveyi (clone 91GP2).  
XX  
PI WO9744361-A1.  
XX  
PD 27-NOV-1997.  
XX  
PF 22-MAY-1997; 97WO-US08793.  
XX  
PR 22-MAY-1996; 96US-0651572.  
XX  
PA (RECO-) RECOMBINANT BIOCATALYSIS INC.  
XX  
PI Lam DE, Mathur EJ;  
XX

DR WPI; 1998-018435/02.  
DR N-PSDB; AAT94216.  
XX  
XX Endoglucanase(s), preferably from archaeal bacterium, AEPRI 1a -  
PT useful to degrade carboxymethylcellulose and hydrolyse of  
PT beta-1,4-glycosidic bonds in cellulose  
XX  
PS Claim 1; Fig 1X; 164pp; English.  
XX  
CC This protein comprises an endoglucanase of Vibrio harveyi  
CC (clone 91GP2) that is capable of degrading carboxymethylcellulose  
CC and of hydrolysing the beta-1,4-glycosidic bonds in cellulose. It  
CC has homology to an endoglucanase of archaeobacterium AEPRI1a (see  
CC AAW34985). It can be produced from native cells or from recombinant  
CC host cells, especially prokaryotic host cells transformed with a  
CC plasmid or virus-derived vector including the endoglucanase DNA  
CC (see AAT94216). 24 Endoglucanases (see AAW34986-W35008) are claimed.  
CC They can be used to degrade cellulose for the conversion of plant  
CC biomass into fuels and chemicals, for use in detergents, textiles,  
CC animal feed, waste treatment, and in the fruit juice and brewing  
CC industries for the clarification and extraction of juices.  
XX  
SQ Sequence 504 AA;  
XX  
Query Match 40.2%; Score 41; DB 19; Length 504;  
Best Local Similarity 46.2%; Pred. No. 1.7e+02;  
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
QY 5 PCGAVMILYFHKW 17  
| | | | : | | | |  
Db 34 PDGVTGVGYWHNW 46  
XX  
RESULT 12  
AAW59445  
ID AAW59445 standard; Protein; 544 AA.  
XX  
AC AAW59445;  
XX  
XX 13-OCT-1998 (first entry)  
XX  
DT Hordeum vulgare MLO protein homologue.  
XX  
DE MLO; mildew; pathogen; resistance.  
XX  
KM Hordeum vulgare.  
XX  
OS Hordeum vulgare.  
XX  
PN WO9804586-A2.  
XX  
PD 05-FEB-1998.  
XX  
PF 29-JUL-1997; 97WO-GB02046.  
XX  
PR 07-MAR-1997; 97GB-0004789.  
PR 29-JUL-1996; 96GB-0015879.  
PR 30-OCT-1996; 96GB-0022626.  
XX  
PA (INNE-) INNES CENT INNOVATIONS LTD JOHN.  
XX  
PI Bueschges R, Panstruga R, Schulzelefer PMJ;  
XX  
DR WPI; 1998-159149/14.  
DR N-PSDB; AAV35030.  
XX  
PT New isolated MLO gene of barley - used to develop products for the  
PT production of transgenic plants which have increased pathogen  
PT resistance  
XX  
PS Claim 8; Fig 14; 150pp; English.  
XX  
CC The sequence is that encoded by a homologue of the MLO gene, wild-type  
CC MLO exerts a negative regulatory function on a pathogen defence response,  
CC such that mutants exhibit a defence response in the absence of pathogen.

CC Down-regulation or out-competition of Mio function may be used to  
CC stimulate a defence response in transgenic plants conferring increased  
CC pathogen resistance, especially resistance to powdery mildew or rust.  
CC The product can also be used for identifying compounds able to  
CC stimulate a defence response in a plant by interaction with  
CC encoded polypeptide.  
XX  
SQ Sequence 544 AA;  
Query Match 40.2%; Score 41; DB 19; Length 544;  
Best Local Similarity 40.0%; Pred. No. 1.8e+02;  
Matches 10; Conservative 1; Mismatches 0; Indels 14; Gaps 1;  
QY 8 AVML-----YFHKWR 18  
| | | | | : | | | | |  
DB 23 AVMLVSVAMEHALHKLGHWFHKWR 47  
RESULT 13  
AAU42094  
ID AAU42094 standard; Protein; 562 AA.  
XX  
AC AAU42094;  
XX  
DT 27-FEB-2002 (first entry)  
XX  
DE Propionibacterium acnes immunogenic protein #2990.  
XX  
KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
KW dermatological; osteopathic; neuroprotectant.  
XX  
OS Propionibacterium acnes.  
XX  
PN WO200181581-A2.  
XX  
PD 01-NOV-2001.  
XX  
PF 20-APR-2001; 2001WO-US12865.  
XX  
PR 21-APR-2000; 2000US-199047P.  
PR 02-JUN-2000; 2000US-208841P.  
PR 07-JUL-2000; 2000US-216747P.  
XX  
XX (CORI-) CORIXA CORP.  
PI Skelky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;  
XX  
DR WPI: 2001-616774/71.  
DR N-PSDB; AASS9516.  
XX  
XX  
PT Propionibacterium acnes polypeptides and nucleic acids useful for  
PT vaccinating against and diagnosing infections, especially useful for  
PT treating acne vulgaris -  
XX  
PS Example 1; SEQ ID NO 3289; 1069pp; English.  
XX  
CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
CC polypeptides. The proteins and their associated DNA sequences are used in  
CC the treatment, prevention and diagnosis of medical conditions caused by  
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
CC P. acnes is also involved in infections of bone, joints and the central  
CC nervous system, however it is particularly involved in the inflammatory  
CC lesions associated with acne vulgaris. A method for detecting the  
CC presence or absence of P. acnes in a patient comprises contacting a  
CC sample with a binding agent that binds to the proteins of the invention  
CC and determining the amount of bound protein in the sample. The  
CC polypeptides may be used as antigens in the production of antibodies  
CC specific for P. acnes proteins. These antibodies can be used to  
CC downregulate expression and activity of P. acnes polypeptides and

CC therefore treat P. acnes infections. The antibodies may also be used as  
CC diagnostic agents for determining P. acnes presence, for example, by  
CC enzyme linked immunosorbent assay (ELISA).  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 562 AA;  
Query Match 40.2%; Score 41; DB 22; Length 562;  
Best Local Similarity 31.8%; Pred. No. 1.9e+02;  
Matches 7; Conservative 4; Mismatches 5; Indels 6; Gaps 1;  
QY 2 TSZPCGAVMI-----LYFHKW 17  
| | | | | : | | | | |  
DB 23 THGPAGAILVCCIPKAMYHNSW 44  
RESULT 14  
ABBA8756  
ID ABBA8756 standard; Protein; 575 AA.  
XX  
AC ABBA8756;  
XX  
DT 05-FEB-2002 (first entry)  
XX  
DE Listeria monocytogenes protein #1460.  
XX  
KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;  
KW vitamin B12; bacterial infection; disease.  
XX  
OS Listeria monocytogenes.  
XX  
PN WO200177335-A2.  
XX  
PD 18-OCT-2001.  
XX  
PF 11-APR-2001; 2001WO-FR01118.  
XX  
PR 11-APR-2000; 2000FR-0004629.  
XX  
PA (INSP ) INST PASTEUR.  
PI Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Dehoux P;  
PI Dussurget O, Chetouni F, Nedjari H, Glaser P, Kunst F, Cossart P;  
PI Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;  
PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;  
PI Chakraborty T, Dommann E, Hain T, Berche P, Charbit A, Durant L;  
PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;  
PI Maduenio E, De Pablos B, Wehlund J, Kaerst U, Entian K, Hauf J;  
PI Rose M, Voss H;  
XX  
DR WPI: 2002-010914/01.  
XX  
XX  
PT Genomic sequence for Listeria monocytogenes, useful e.g. for treatment  
PT and prevention of Listeria and related bacterial infections, and  
PT related polypeptides -  
XX  
PS Claim 5; SEQ ID NO 1461; 192pp; French.  
XX  
CC The present invention relates to the genome sequence of Listeria  
CC monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of  
CC it are useful for selecting probes and primers for detecting genes in L.  
CC monocytogenes and related organisms, and for studying genetic  
CC polymorphisms and other genomes. The present sequence is a protein  
CC encoded by the genome sequence of the present invention. Proteins  
CC expressed from the genome sequence are useful for raising specific  
CC antibodies, identification of L. monocytogenes and related organisms, and  
CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin  
CC B12. The genome sequence and proteins encoded by it are also useful for  
CC selecting compounds that regulate gene expression and cell replication  
CC and modulate L. monocytogenes-related diseases. In addition, the genome  
CC sequence and proteins encoded by it are useful in pharmaceutical and

CC vaccines compositions for the treatment or prevention of infections by L.  
CC monocytogenes and related organisms.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 575 AA;

QY 7 GAVMILYFHKWR 18  
11::11:1 1:  
Db 150 GAIILFFMDWK 161

Query Match 40.2%; Score 41; DB 23; Length 575;  
Best Local Similarity 50.0%; Pred. No. 1.9e+02;  
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

RESULT 15  
AAY08613  
ID AAY08613 standard; Protein; 60 AA.

XX AC AAY08613;

XX 05-AUG-1999 (first entry)

DE Inter-alpha-trypsin inhibitor (BPI type) protein fragment TIBOR.

XX APP; beta-amyloid precursor protein; human; transgenic mice; pathology;  
KW Alzheimer's disease; model; therapeutic compound; brain; mechanism;  
KW nerve tissue specific promoter; synthesis; inhibitor; deposition;  
KW plaque formation; treatment.

XX Unidentified.

OS US5912410-A.

XX 15-JUN-1999.

XX PF 13-APR-1995; 95US-0422333.

XX PR 21-OCT-1994; 94US-0327381.

XX PR 15-JUN-1990; 90US-0538857.

XX PR 17-JUN-1991; 91US-0716725.

XX PR 13-APR-1995; 95US-0422333.

XX (SCIO-) SCIOS INC.

XX Cordell B;

XX WPI; 1999-357231/30.

PS Transgenic mice useful for studying compounds potentially useful in  
the treatment of Alzheimer's disease

XX Disclosure; Fig 8B; 72pp; English.  
CC This invention describes novel transgenic mice expressing proteins  
CC related to the pathology of Alzheimer's disease and which provide models  
CC for studying potentially therapeutic compounds. The transgenic mice  
CC contain a DNA sequence encoding a beta-amyloid precursor protein (APP)  
CC and a nerve tissue specific promoter operably linked to the beta-APP  
CC allowing its expression to form beta-amyloid protein deposits in the  
CC animal's brain. The transgenic mouse is useful for elucidating the  
CC molecular mechanisms involved in the synthesis of and, more importantly,  
CC inhibiting the synthesis and deposition of beta-amyloid proteins (most  
CC importantly in the brain where plaque formation is associated with  
CC Alzheimer's disease) by inhibiting production and/or increasing cleavage  
CC after production. The transgenic animals provide useful models for  
CC studying the in vivo relationships of the proteins to each other and to  
CC other compounds being tested for their usefulness in treating Alzheimer's  
CC disease.

XX Sequence 60 AA;

Query Match 39.2%; Score 40; DB 20; Length 60;  
Best Local Similarity 63.6%; Pred. No. 26;  
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 5 PCGAVMILYFH 15  
11111111:  
Db 15 PCKAAMIRYFY 25

Search completed: November 15, 2002, 11:44:12  
Job time : 36 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 15, 2002, 11:43:15 ; Search time 14 Seconds  
(without alignments)  
37.829 Million cell updates/sec

Title: US-09-881-569A-2  
Perfect score: 102  
Sequence: 1 B7SZPCGAVMILYFKWR 18

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Minimum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*

- 1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep:\*
- 2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep:\*
- 3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep:\*
- 4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep:\*
- 5: /cgn2\_6/ptodata/1/1aa/PCTUS\_COMB.pep:\*
- 6: /cgn2\_6/ptodata/1/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID                  | Description        |
|------------|-------|-------------|--------|----|---------------------|--------------------|
| 1          | 40    | 39.2        | 60     | 1  | US-08-358-160-112   | Sequence 112, App  |
| 2          | 40    | 39.2        | 60     | 2  | US-08-422-333-14    | Sequence 14, Appl  |
| 3          | 40    | 39.2        | 60     | 6  | 5187153-22          | Patent No. 5187153 |
| 4          | 40    | 39.2        | 60     | 6  | 5220013-25          | Patent No. 5220013 |
| 5          | 40    | 39.2        | 60     | 6  | 5223482-24          | Patent No. 5223482 |
| 6          | 40    | 39.2        | 60     | 6  | 5466783-5           | Patent No. 5466783 |
| 7          | 40    | 39.2        | 476    | 4  | US-09-632-021-2     | Sequence 2, Appl1  |
| 8          | 39    | 38.2        | 111    | 1  | US-08-084-718-61    | Sequence 61, Appl1 |
| 9          | 39    | 38.2        | 111    | 1  | US-08-443-976-61    | Sequence 61, Appl1 |
| 10         | 39    | 38.2        | 111    | 1  | US-08-443-977-61    | Sequence 61, Appl1 |
| 11         | 39    | 38.2        | 504    | 4  | US-09-499-302A-9    | Sequence 9, Appl1  |
| 12         | 38    | 37.3        | 69     | 1  | US-08-358-160-115   | Sequence 115, App  |
| 13         | 37.5  | 36.8        | 346    | 4  | US-09-134-001C-5196 | Sequence 5196, Ap  |
| 14         | 37    | 36.3        | 58     | 1  | US-08-358-160-110   | Sequence 110, App  |
| 15         | 37    | 36.3        | 59     | 1  | US-08-358-160-111   | Sequence 111, App  |
| 16         | 37    | 36.3        | 323    | 4  | US-08-878-474-7     | Sequence 7, Appl1  |
| 17         | 37    | 36.3        | 325    | 4  | US-08-878-474-9     | Sequence 9, Appl1  |
| 18         | 37    | 36.3        | 394    | 4  | US-08-791-115B-27   | Sequence 27, Appl1 |
| 19         | 37    | 36.3        | 403    | 4  | US-08-791-115B-1    | Sequence 1, Appl1  |
| 20         | 37    | 36.3        | 403    | 4  | US-08-791-115B-23   | Sequence 23, Appl1 |
| 21         | 37    | 36.3        | 430    | 4  | US-08-791-115B-25   | Sequence 25, Appl1 |
| 22         | 37    | 36.3        | 430    | 4  | US-08-791-115B-7    | Sequence 7, Appl1  |
| 23         | 37    | 36.3        | 536    | 2  | US-08-484-993B-4    | Sequence 4, Appl1  |
| 24         | 37    | 36.3        | 536    | 2  | US-08-484-158B-4    | Sequence 4, Appl1  |
| 25         | 37    | 36.3        | 536    | 2  | US-08-484-596A-4    | Sequence 4, Appl1  |
| 26         | 37    | 36.3        | 536    | 2  | US-08-480-150A-4    | Sequence 4, Appl1  |
| 27         | 37    | 36.3        | 536    | 3  | US-08-458-731-4     | Sequence 4, Appl1  |

|    |      |      |     |   |                    |                    |
|----|------|------|-----|---|--------------------|--------------------|
| 28 | 37   | 36.3 | 536 | 3 | US-08-149-223A-4   | Sequence 4, Appl1  |
| 29 | 37   | 36.3 | 565 | 4 | US-08-906-156A-12  | Sequence 12, Appl1 |
| 30 | 37   | 36.3 | 645 | 4 | US-08-791-115B-6   | Sequence 6, Appl1  |
| 31 | 37   | 36.3 | 742 | 4 | US-08-791-115B-5   | Sequence 5, Appl1  |
| 32 | 37   | 36.3 | 900 | 4 | US-08-890-865A-4   | Sequence 4, Appl1  |
| 33 | 36.5 | 35.8 | 238 | 4 | US-08-944-483-39   | Sequence 39, Appl1 |
| 34 | 36.5 | 35.8 | 262 | 1 | US-08-744-026-4    | Sequence 4, Appl1  |
| 35 | 36.5 | 35.8 | 262 | 2 | US-08-790-137-1    | Sequence 1, Appl1  |
| 36 | 36.5 | 35.8 | 262 | 2 | US-08-790-137-3    | Sequence 3, Appl1  |
| 37 | 36.5 | 35.8 | 262 | 2 | US-08-681-151-4    | Sequence 4, Appl1  |
| 38 | 36.5 | 35.8 | 262 | 2 | US-09-102-732-4    | Sequence 4, Appl1  |
| 39 | 36.5 | 35.8 | 262 | 2 | US-08-824-874-4    | Sequence 4, Appl1  |
| 40 | 36.5 | 35.8 | 262 | 3 | US-08-807-151-4    | Sequence 4, Appl1  |
| 41 | 36.5 | 35.8 | 262 | 4 | US-09-261-767-4    | Sequence 4, Appl1  |
| 42 | 36.5 | 35.8 | 262 | 4 | US-09-210-084-4    | Sequence 4, Appl1  |
| 43 | 36.5 | 35.8 | 262 | 4 | US-09-478-957-4    | Sequence 4, Appl1  |
| 44 | 36   | 35.3 | 15  | 1 | US-07-598-737C-17  | Sequence 17, Appl1 |
| 45 | 36   | 35.3 | 19  | 4 | US-08-974-549A-118 | Sequence 118, App  |

## ALIGNMENTS

RESULT 1  
US-08-358-160-112  
; Sequence 112, Application US/08358160  
; Patent No. 5663143  
; GENERAL INFORMATION:  
; APPLICANT: LEY, Arthur C.  
; APPLICANT: LADNER, Robert C.  
; APPLICANT: GUTERMAN, Sonia K.  
; APPLICANT: ROBERTS, Bruce L.  
; APPLICANT: MARKLAND, William  
; APPLICANT: KENT, Rachel B.  
; TITLE OF INVENTION: ENGINEERED HUMAN-DERIVED KUNITZ  
; TITLE OF INVENTION: DOMAINS THAT INHIBIT HUMAN NEUTROPHIL ELASTASE  
; NUMBER OF SEQUENCES: 234  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROWDY AND NEIMARK  
; STREET: 419 Seventh Street, N.W. Suite 300  
; CITY: Washington  
; STATE: District of Columbia  
; COUNTRY: USA  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/358,160  
; FILING DATE: 16-DEC-1994  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/133,031  
; FILING DATE: 13-OCT-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/009,319  
; FILING DATE: 26-JAN-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/664,989  
; FILING DATE: 01-MAR-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/487,063  
; FILING DATE: 02-MAR-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/240,160  
; FILING DATE: 02-SEP-1988  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cooper, Iver P.  
; REGISTRATION NUMBER: 28,005  
; REFERENCE/DOCKET NUMBER: LEY-1  
; TELECOMMUNICATION INFORMATION:





```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/361,912
; FILING DATE: 06-JUN-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 359,911
; FILING DATE: 12-MAY-1989
; APPLICATION NUMBER: 87,002
; FILING DATE: 18-AUG-1987
; APPLICATION NUMBER: 8,810
; FILING DATE: 30-JAN-1987
; APPLICATION NUMBER: 948,376
; FILING DATE: 31-DEC-1986
; APPLICATION NUMBER: 932,193
; FILING DATE: 17-NOV-1986
; SEQ ID NO:24:
; LENGTH: 60
5223482-24

Query Match      39.2%; Score 40; DB 6; Length 60;
Best Local Similarity 63.6%; Pred. No. 35;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY      5 PCGAVMILYFH 15
        ||| ||| |||
        15 PCKAMIRYFY 25

RESULT 6
5466783-5
; Patent No. 5466783
; APPLICANT: Wan, Tze-Cheln.;Kretzmer, Kunlko K.;Broze,
; George J. Jr.
; TITLE OF INVENTION: HUMAN TISSUE FACTOR INHIBITOR
; NUMBER OF SEQUENCES: 26
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/93,285
; FILING DATE: 15-JUL-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 566,280
; FILING DATE: 13-AUG-1990
; APPLICATION NUMBER: 123,753
; FILING DATE: 23-NOV-1987
; APPLICATION NUMBER: 77,366
; FILING DATE: 23-JUL-1987
; SEQ ID NO:5:
; LENGTH: 60
5466783-5

Query Match      39.2%; Score 40; DB 6; Length 60;
Best Local Similarity 63.6%; Pred. No. 35;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

5 PCGAVMILYFH 15
||| ||| |||
15 PCKAMIRYFY 25

Db      15 PCKAMIRYFY 25

RESULT 7
US-09-632-021-2
; Sequence 2, Application US/09632021
; Patent No. 6406896
; GENERAL INFORMATION:
; APPLICANT: Reznikoff, William S
; APPLICANT: Naumann, Todd A
; TITLE OF INVENTION: Transposase Enzyme and Method for use
; FILE REFERENCE: 960296.96471
; CURRENT APPLICATION NUMBER: US/09/632,021
; CURRENT FILING DATE: 2000-08-02
; PRIOR APPLICATION NUMBER: 60/146686
; PRIOR FILING DATE: 1999-08-02
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 2
```

```

; LENGTH: 476
; TYPE: PRT
; ORGANISM: Transposon Tns
; US-09-632-021-2

Query Match      39.2%; Score 40; DB 4; Length 476;
Best Local Similarity 34.8%; Pred. No. 2.2e+02;
Matches 8; Conservative 5; Mismatches 4; Indels 6; Gaps 1;

QY      2 TS2PCGA-----VMILYFHKWR 18
        ||| | : | : | : | : | : |
        302 TSEPVESTLAQALRVIDYTHRW 324

Db      302 TSEPVESTLAQALRVIDYTHRW 324

RESULT 8
US-08-084-718-61
; Sequence 61, Application US/08084718
; Patent No. 5591603
; GENERAL INFORMATION:
; APPLICANT: Bjorn, Soeren E.
; APPLICANT: No. 5591603r1s, Kjeld
; APPLICANT: Diness, V1990
; APPLICANT: No. 5591603rskov-Lauritsen, Lelf
; APPLICANT: Christensen, Niels D.
; APPLICANT: Bregengaard, Claus
; APPLICANT: No. 5591603r1s, Fanny
; APPLICANT: Petersen, Lars C.
; TITLE OF INVENTION: Process for the Preparation of Aprotinin
; NUMBER OF SEQUENCES: 83
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NO. 55916030 No. 5591603disk of No. 5591603th America, Inc.
; STREET: 405 Lexington Avenue, 62nd Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/084,718
; FILING DATE: 19930623
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 4501/87
; FILING DATE: 28-AUG-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 2254/88
; FILING DATE: 26-APR-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK88/00138
; FILING DATE: 28-AUG-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK89/00096
; FILING DATE: 25-APR-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/466,408
; FILING DATE: 21-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 2361/90
; FILING DATE: 01-OCT-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/598,337
; FILING DATE: 19-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 1118/91
; FILING DATE: 12-JUN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK91/00299
; FILING DATE: 01-OCT-1991
```

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/827,687  
FILING DATE: 29-JAN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/024,925  
FILING DATE: 26-FEB-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Agtis, Cheryl H.  
REGISTRATION NUMBER: 34,086  
REFERENCE/DOCKET NUMBER: 3143.224-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0123  
TELEFAX: 212-867-0298  
INFORMATION FOR SEQ ID NO: 61:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 111 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-084-718-61

Query Match 38.2%; Score 39; DB 1; Length 111;  
Best Local Similarity 63.6%; Pred. No. 84;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 5 PCGAVMILYFH 15  
|||:||||:  
DB 66 PCKARIILYFY 76

RESULT 9  
US-08-443-976-61  
Sequence 61, Application US/08443976  
Patent No. 5618915  
GENERAL INFORMATION:  
APPLICANT: Bjorn, Soeren E.  
APPLICANT: No. 5618915r1s, Kjeld  
APPLICANT: Diness, Viggo  
APPLICANT: No. 5618915rskov-Lauritsen, Lelf  
APPLICANT: Christensen, Niels D.  
APPLICANT: Bregengaard, Claus  
APPLICANT: No. 5618915r1s, Fanny  
APPLICANT: Petersen, Lars C.  
TITLE OF INVENTION: Process for the Preparation of Aprotinin  
TITLE OF INVENTION: and Aprotinin Analogs  
NUMBER OF SEQUENCES: 83  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 5618915o No. 5618915disk of No. 5618915th America, Inc.  
STREET: 405 Lexington Avenue, 62nd Floor  
CITY: New York  
STATE: New York  
COUNTRY: United States of America  
ZIP: 10174-6201  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/443,976  
FILING DATE: 18-MAY-1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/084,718  
FILING DATE: 23-JUN-1993  
APPLICATION NUMBER: DK 4501/87  
FILING DATE: 28-AUG-1987  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DK 2254/88  
FILING DATE: 26-APR-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/DK88/00138  
FILING DATE: 28-AUG-1988

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/DK89/00096  
FILING DATE: 25-APR-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/466,408  
FILING DATE: 21-JUN-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DK 2361/90  
FILING DATE: 01-OCT-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/598,337  
FILING DATE: 19-NOV-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DK 1118/91  
FILING DATE: 12-JUN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/DK91/00299  
FILING DATE: 01-OCT-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/827,687  
FILING DATE: 29-JAN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/024,925  
FILING DATE: 26-FEB-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Agtis, Cheryl H.  
REGISTRATION NUMBER: 34,086  
REFERENCE/DOCKET NUMBER: 3143.224-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0123  
TELEFAX: 212-867-0298  
INFORMATION FOR SEQ ID NO: 61:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 111 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-443-976-61

Query Match 38.2%; Score 39; DB 1; Length 111;  
Best Local Similarity 63.6%; Pred. No. 84;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 5 PCGAVMILYFH 15  
|||:||||:  
DB 66 PCKARIILYFY 76

RESULT 10  
US-08-443-977-61  
Sequence 61, Application US/08443977  
Patent No. 5621074  
GENERAL INFORMATION:  
APPLICANT: Bjorn, Soeren E.  
APPLICANT: No. 5621074r1s, Kjeld  
APPLICANT: Diness, Viggo  
APPLICANT: No. 5621074rskov-Lauritsen, Lelf  
APPLICANT: Christensen, Niels D.  
APPLICANT: Bregengaard, Claus  
APPLICANT: No. 5621074r1s, Fanny  
APPLICANT: Petersen, Lars C.  
TITLE OF INVENTION: Process for the Preparation of Aprotinin  
TITLE OF INVENTION: and Aprotinin Analogs  
NUMBER OF SEQUENCES: 83  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 5621074o No. 5621074disk of No. 5621074th America, Inc.  
STREET: 405 Lexington Avenue, 62nd Floor  
CITY: New York  
STATE: New York  
COUNTRY: United States of America  
ZIP: 10174-6201  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

```
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/443,977
; FILING DATE: 18-MAY-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/084,718
; FILING DATE: 23-JUN-1993
; APPLICATION NUMBER: DK 4501/87
; FILING DATE: 28-AUG-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 2254/88
; FILING DATE: 26-APR-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK88/00138
; FILING DATE: 28-AUG-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK89/00096
; FILING DATE: 25-APR-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/466,408
; FILING DATE: 21-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 2361/90
; FILING DATE: 01-OCT-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/598,337
; FILING DATE: 19-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 1118/91
; FILING DATE: 12-JUN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK91/00299
; FILING DATE: 01-OCT-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/827,687
; FILING DATE: 29-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/024,925
; FILING DATE: 26-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Agtis, Cheryl H.
; REGISTRATION NUMBER: 34,086
; REFERENCE/DOCKET NUMBER: 3143.224-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-867-0298
; INFORMATION FOR SEQ ID NO: 61:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 111 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-443-977-61

Query Match      38.2%; Score 39; DB 1; Length 111;
Best Local Similarity 63.6%; Pred. No. 84;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
```

OY 5 PCGAVMILYFH 15  
| | | : | | | :  
Db 66 PCKARILYFY 76

```
RESULT 11
US-09-499-302A-9
; Sequence 9, Application US/09499302A
; Patent No. 6369212
; GENERAL INFORMATION:
; APPLICANT: BOUNG-JUN, OH
; APPLICANT: MOON, KYUNG KO
```

```
; APPLICANT: YOUNG, SOON KIM
; TITLE OF INVENTION: A CYTOCHROME P450 GENE HIGHLY EXPRESSED IN THE
; TITLE OF INVENTION: INCOMPATIBLE INTERACTION
; FILE REFERENCE: 10324/P64443USO
; CURRENT APPLICATION NUMBER: US/09/499,302A
; CURRENT FILING DATE: 2000-02-07
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 9
; LENGTH: 504
; TYPE: PRT
; ORGANISM: Arabidopsis sp.
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (1)
; OTHER INFORMATION: variable or unknown amino acid
; NAME/KEY: MOD_RES
; LOCATION: (62)
; OTHER INFORMATION: variable or unknown amino acid
; NAME/KEY: MOD_RES
; LOCATION: (243)
; OTHER INFORMATION: variable or unknown amino acid
; NAME/KEY: MOD_RES
; LOCATION: (246)
; OTHER INFORMATION: variable or unknown amino acid
; NAME/KEY: MOD_RES
; LOCATION: (253)
; OTHER INFORMATION: variable or unknown amino acid
; NAME/KEY: MOD_RES
; LOCATION: (324)
; OTHER INFORMATION: variable or unknown amino acid
; NAME/KEY: MOD_RES
; LOCATION: (394)
; OTHER INFORMATION: variable or unknown amino acid
; US-09-499-302A-9
```

Query Match 38.2%; Score 39; DB 4; Length 504;  
Best Local Similarity 54.5%; Pred. No. 3.2e+02;  
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 7 GAVMILYFHKW 17  
| : | : | | | | |  
Db 459 GLMLILYFDM 469

```
RESULT 12
US-08-358-160-115
; Sequence 115, Application US/08358160
; Patent No. 5663143
; GENERAL INFORMATION:
; APPLICANT: LEY, Arthur C.
; APPLICANT: LADNER, Robert C.
; APPLICANT: GUTERMAN, Sonia K.
; APPLICANT: ROBERTS, Bruce L.
; APPLICANT: MARKLAND, William
; APPLICANT: KENT, Rachel B.
; TITLE OF INVENTION: ENGINEERED HUMAN-DERIVED KUNITZ
; TITLE OF INVENTION: DOMAINS THAT INHIBIT HUMAN NEUTROPHIL ELASTASE
; NUMBER OF SEQUENCES: 234
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W. Suite 300
; CITY: Washington
; STATE: District of Columbia
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/358,160
```

FILING DATE: 16-DEC-1994  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/133,031  
FILING DATE: 13-OCT-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/009,319  
FILING DATE: 26-JAN-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/664,989  
FILING DATE: 01-MAR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/487,063  
FILING DATE: 02-MAR-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/240,160  
FILING DATE: 02-SEP-1988  
ATTORNEY/AGENT INFORMATION:  
NAME: Cooper, Iver P.  
REGISTRATION NUMBER: 28,005  
REFERENCE/DOCKET NUMBER: LEY-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
TELEX: 248633  
INFORMATION FOR SEQ ID NO: 115:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 69 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-358-160-115

Query Match 37.3%; Score 38; DB 1; Length 69;  
Best Local Similarity 63.6%; Pred. No. 77;  
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 5 PCGAVMILYFH 15  
||| ||| |||  
Db 18 PCRAKMRIFY 28

## RESULT 13

-09-134-001C-5196  
Sequence 5196, Application US/09134001C  
Patent No. 6380370  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucelte-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: GTC-007  
CURRENT APPLICATION NUMBER: US/09/134,001C  
FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/064,964  
PRIOR FILING DATE: 1997-11-08  
PRIOR APPLICATION NUMBER: US 60/055,779  
PRIOR FILING DATE: 1997-08-14  
NUMBER OF SEQ ID NOS: 5674  
SEQ ID NO 5196  
LENGTH: 346  
TYPE: PRT  
ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-5196

Query Match 36.8%; Score 37.5; DB 4; Length 346;  
Best Local Similarity 43.8%; Pred. No. 3.7e+02;  
Matches 7; Conservative 3; Mismatches 5; Indels 1; Gaps 1;

OY 3 SZPCGAVMILYFH-K-W 17  
||| ||| ||| |||  
Db 129 SIPLGITAAVYHKIM 144

RESULT 14  
US-08-358-160-110  
Sequence 110, Application US/08358160  
Patent No. 5653143

GENERAL INFORMATION:  
APPLICANT: Ley, Arthur C.  
APPLICANT: LADNER, Robert C.  
APPLICANT: GUTERMAN, Sonia K.  
APPLICANT: ROBERTS, Bruce L.  
APPLICANT: MARKLAND, William  
APPLICANT: KENT, Rachel B.  
TITLE OF INVENTION: ENGINEERED HUMAN-DERIVED KUNITZ  
TITLE OF INVENTION: DOMAINS THAT INHIBIT HUMAN NEUTROPHIL ELASTASE  
NUMBER OF SEQUENCES: 234  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BROWDY AND NEIMARK  
STREET: 419 Seventh Street, N.W. Suite 300  
CITY: Washington  
STATE: District of Columbia  
COUNTRY: USA  
ZIP: 20004

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/358,160  
FILING DATE: 16-DEC-1994  
CLASSIFICATION: 514

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/133,031  
FILING DATE: 13-OCT-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/009,319  
FILING DATE: 26-JAN-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/664,989  
FILING DATE: 01-MAR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/487,063  
FILING DATE: 02-MAR-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/240,160  
FILING DATE: 02-SEP-1988  
ATTORNEY/AGENT INFORMATION:  
NAME: Cooper, Iver P.  
REGISTRATION NUMBER: 28,005  
REFERENCE/DOCKET NUMBER: LEY-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
TELEX: 248633

INFORMATION FOR SEQ ID NO: 110:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 58 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-358-160-110

Query Match 36.3%; Score 37; DB 1; Length 58;  
Best Local Similarity 63.6%; Pred. No. 91;  
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 5 PCGAVMILYFH 15  
||| ||| |||  
Db 13 PCRAKMRIFY 23

## RESULT 15

```

US-08-358-160-111
; Sequence 111, Application US/08358160
; Patent No. 5663143
; GENERAL INFORMATION:
; APPLICANT: LEY, Arthur C.
; APPLICANT: LADNER, Robert C.
; APPLICANT: GUTERMAN, Sonia K.
; APPLICANT: ROBERTS, Bruce L.
; APPLICANT: MARKLAND, William
; APPLICANT: KENT, Rachel B.
; TITLE OF INVENTION: ENGINEERED HUMAN-DERIVED KUNITZ
; TITLE OF INVENTION: DOMAINS THAT INHIBIT HUMAN NEUTROPHIL ELASTASE
; NUMBER OF SEQUENCES: 234
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W. Suite 300
; CITY: Washington
; STATE: District of Columbia
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/358,160
; FILING DATE: 16-DEC-1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/133,031
; FILING DATE: 13-OCT-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/009,319
; FILING DATE: 26-JAN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/664,989
; FILING DATE: 01-MAR-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/487,063
; FILING DATE: 02-MAR-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/240,160
; FILING DATE: 02-SEP-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Cooper, Iver P.
; REGISTRATION NUMBER: 28,005
; REFERENCE/DOCKET NUMBER: LEY-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 111:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 59 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-358-160-111

Query Match          36.3%; Score 37; DB 1; Length 59;
Best Local Similarity 63.6%; Pred.No. 92;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY      5 PCGAVMILYFH 15
      ||| |||
Db      14 PCKARMIRYFY 24

```

Search completed: November 15, 2002, 11:45:31  
 Job time : 15 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: November 15, 2002, 11:44:39 ; Search time 10 Seconds  
(Without alignments)  
27.110 Million cell updates/sec

Title: US-09-881-569A-2  
Perfect score: 102  
Sequence: 1 BTRSPCGAVMILYFHKMR 18

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 97044 seqs, 15060890 residues

Total number of hits satisfying chosen parameters: 97044

Minimum DB seq length: 0

Minimum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep:\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*
- 6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*
- 7: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep:\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep:\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep:\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep:\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description         |
|------------|-------|-------------|--------|-------|---------------------|
| 1          | 102   | 100.0       | 18     | 10    | US-09-881-569A-2    |
| 2          | 59    | 57.8        | 16     | 10    | US-09-881-569A-1    |
| 3          | 44    | 43.1        | 893    | 10    | US-09-906-453-4     |
| 4          | 39    | 38.2        | 39     | 10    | US-09-864-761-48038 |
| 5          | 38    | 37.3        | 74     | 10    | US-09-764-870-484   |
| 6          | 38    | 37.3        | 1185   | 9     | US-09-895-913A-246  |
| 7          | 38    | 37.3        | 1185   | 10    | US-09-815-242-11466 |
| 8          | 37    | 36.3        | 111    | 10    | US-09-864-761-37111 |
| 9          | 37    | 36.3        | 318    | 12    | US-10-028-051-8     |
| 10         | 37    | 36.3        | 323    | 9     | US-09-903-170C-7    |
| 11         | 37    | 36.3        | 323    | 10    | US-09-903-180B-7    |
| 12         | 37    | 36.3        | 323    | 10    | US-09-903-187A-7    |
| 13         | 37    | 36.3        | 323    | 10    | US-09-903-171A-7    |
| 14         | 37    | 36.3        | 323    | 10    | US-09-903-188A-7    |
| 15         | 37    | 36.3        | 323    | 10    | US-09-903-323A-7    |
| 16         | 37    | 36.3        | 323    | 10    | US-09-903-325A-7    |
| 17         | 37    | 36.3        | 325    | 9     | US-09-903-170C-9    |
| 18         | 37    | 36.3        | 325    | 9     | US-09-974-298-17    |
| 19         | 37    | 36.3        | 325    | 10    | US-09-903-180B-9    |

|    |      |      |      |    |                     |                    |
|----|------|------|------|----|---------------------|--------------------|
| 20 | 37   | 36.3 | 325  | 10 | US-09-903-187A-9    | Sequence 9, Appli  |
| 21 | 37   | 36.3 | 325  | 10 | US-09-903-171A-9    | Sequence 9, Appli  |
| 22 | 37   | 36.3 | 325  | 10 | US-09-903-188A-9    | Sequence 9, Appli  |
| 23 | 37   | 36.3 | 325  | 10 | US-09-903-323A-9    | Sequence 9, Appli  |
| 24 | 37   | 36.3 | 325  | 10 | US-09-903-325A-9    | Sequence 9, Appli  |
| 25 | 37   | 36.3 | 325  | 12 | US-10-028-051-2     | Sequence 2, Appli  |
| 26 | 37   | 36.3 | 325  | 12 | US-10-028-051-4     | Sequence 4, Appli  |
| 27 | 37   | 36.3 | 325  | 12 | US-10-028-051-2     | Sequence 2, Appli  |
| 28 | 37   | 36.3 | 403  | 10 | US-09-870-379-2     | Sequence 194, App  |
| 29 | 36.5 | 35.8 | 71   | 12 | US-10-001-879-194   | Sequence 9, Appli  |
| 30 | 36.5 | 35.8 | 262  | 10 | US-09-796-294-9     | Sequence 4, Appli  |
| 31 | 36.5 | 35.8 | 262  | 10 | US-09-764-762-4     | Sequence 7, Appli  |
| 32 | 36.5 | 35.8 | 262  | 10 | US-09-988-975A-7    | Sequence 745, App  |
| 33 | 36.5 | 35.8 | 267  | 10 | US-09-925-297-745   | Sequence 44551, A  |
| 34 | 36.5 | 35.3 | 28   | 10 | US-09-864-761-44551 | Sequence 204, App  |
| 35 | 36.5 | 35.3 | 49   | 10 | US-08-896-322-5     | Sequence 5, Appli  |
| 36 | 36.5 | 35.3 | 57   | 8  | US-09-764-887-204   | Sequence 1624, Ap  |
| 37 | 36.5 | 35.3 | 276  | 10 | US-09-925-300-1624  | Sequence 4, Appli  |
| 38 | 36.5 | 35.3 | 552  | 9  | US-09-843-676-4     | Sequence 25, Appli |
| 39 | 36.5 | 35.3 | 639  | 10 | US-09-768-877-25    | Sequence 8, Appli  |
| 40 | 36.5 | 35.3 | 842  | 10 | US-09-798-831-8     | Sequence 45, Appli |
| 41 | 36.5 | 35.3 | 891  | 10 | US-09-949-192-45    | Sequence 306, App  |
| 42 | 36.5 | 35.3 | 1356 | 10 | US-09-801-368-306   | Sequence 44559, A  |
| 43 | 36.5 | 34.3 | 38   | 10 | US-09-864-761-44559 | Sequence 1085, Ap  |
| 44 | 36.5 | 34.3 | 66   | 10 | US-09-764-877-1085  | Sequence 5, Appli  |
| 45 | 36.5 | 34.3 | 136  | 10 | US-09-970-711-5     | Sequence 316, App  |

## ALIGNMENTS

RESULT 1  
US-09-881-569A-2  
; Sequence 2, Application US/09881569A  
; Patent No. US20020151687A1  
GENERAL INFORMATION:  
; APPLICANT: KHANNA, PUSHPA  
; TITLE OF INVENTION: PROTEIN/POLYPEPTIDE-K OBTAINED FROM MOMORDICA CHARANTIA  
; FILE REFERENCE: 013488-3  
; CURRENT APPLICATION NUMBER: US/09/881.569A  
; PRIOR FILING DATE: 2001-06-14  
; PRIOR APPLICATION NUMBER: 560/DEL/99  
; PRIOR FILING DATE: 1999-04-13  
; PRIOR APPLICATION NUMBER: 561/DEL/99  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: Patentln Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 18  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Polypeptide-k  
US-09-881-569A-2  
Query Match  
Best Local Similarity 100.0%; Score 102; DB 10; Length 18;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 BTRSPCGAVMILYFHKMR 18  
Db 1 BTRSPCGAVMILYFHKMR 18  
RESULT 2  
US-09-881-569A-1  
; Sequence 1, Application US/09881569A  
; Patent No. US20020151687A1  
; GENERAL INFORMATION:  
; APPLICANT: KHANNA, PUSHPA  
; TITLE OF INVENTION: PROTEIN/POLYPEPTIDE-K OBTAINED FROM MOMORDICA CHARANTIA

```

; TITLE OF INVENTION: AND A PROCESS FOR THE EXTRACTION THEREOF
; FILE REFERENCE: 013488-3
; CURRENT APPLICATION NUMBER: US/09/881,569A
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: 560/DEL/99
; PRIOR FILING DATE: 1999-04-13
; PRIOR APPLICATION NUMBER: 561/DEL/99
; PRIOR FILING DATE: 1999-04-13
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 16
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Polypeptide-p
-09-881-569A-1

```

```

Query Match      57.8%; Score 59; DB 10; Length 16;
Best Local Similarity 62.5%; Pred. No. 0.0018;
Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

```

```

QY 1 BTSZPCGAVMILYFHK 16
   :||:| | | | | | |
Db 1 DTSEPGAVCMILYFHK 16

```

```

RESULT 3
US-09-906-453-4
; Sequence 4, Application US/09906453
; Patent No. US20020120125A1
; GENERAL INFORMATION:
; APPLICANT: Kaeppler, Shawn
; APPLICANT: Springer, Nathan
; APPLICANT: Helentjaris, Timothy
; APPLICANT: Phillips, Ronald
; TITLE OF INVENTION: Polycomb genes from Maize - Mez1 and Mez2
; FILE REFERENCE: WISCONSIN100US
; CURRENT APPLICATION NUMBER: US/09/906,453
; CURRENT FILING DATE: 2001-07-16
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 4
; LENGTH: 893
; TYPE: PRT
; ORGANISM: Zea mays
US-09-906-453-4

```

```

Query Match      43.1%; Score 44; DB 10; Length 893;
Best Local Similarity 38.9%; Pred. No. 18;
Matches 7; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

```

```

QY 1 BTSZPCGAVMILYFHK 18
   :||:| | | | | | |
Db 317 ENKKPCGHLCTLRFPQWR 334

```

```

RESULT 4
US-09-864-761-48038
; Sequence 48038, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecm1ca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312

```

```

; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 48038
; LENGTH: 39
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC009966.2
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.9
US-09-864-761-48038

```

```

Query Match      38.2%; Score 39; DB 10; Length 39;
Best Local Similarity 46.7%; Pred. No. 5.7;
Matches 7; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

```

```

QY 3 SZPCGAVMILYFHK 17
   |:| | | | | |
Db 21 SEPTGLEMIHYSYLW 35

```

```

RESULT 5
US-09-764-870-484
; Sequence 484, Application US/09764870
; Patent No. US20020042386A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; FILE REFERENCE: PT214
; CURRENT APPLICATION NUMBER: US/09/764,870
; CURRENT FILING DATE: 2001-01-17
; NUMBER OF SEQ ID NOS: 646
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 484
; LENGTH: 74
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-870-484

```



Query Match 37.3%; Score 38; DB 10; Length 74;  
Best Local Similarity 63.6%; Pred. No. 15;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
QY 2 TSZPCGAVMIL 12  
11 11111:  
Db 5 TSRCGAVVII 15

## RESULT 6

US-09-895-913A-246  
; Sequence 246, Application US/09895913A  
; Patent No. US20020160456A1  
; GENERAL INFORMATION:  
; APPLICANT: Kleanthous, Harold  
; APPLICANT: Al-Garawi, Amal  
; APPLICANT: Miller, Charles  
; APPLICANT: Tomb, Jean Francois  
; APPLICANT: Oomen, Raymond P.  
; TITLE OF INVENTION: Identification of Polynucleotides  
; TITLE OF INVENTION: Encoding No. US20020160456A1 Helicobacter Polypeptides in the  
; TITLE OF INVENTION: Genome  
; FILE REFERENCE: 06132/043002  
; CURRENT APPLICATION NUMBER: US/09/895,913A  
; CURRENT FILING DATE: 2001-06-29  
; PRIOR APPLICATION NUMBER: US 08/881,227  
; PRIOR FILING DATE: 1997-06-24  
; NUMBER OF SEQ ID NOS: 368  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 246  
; LENGTH: 1185  
; TYPE: PRT  
; ORGANISM: Helicobacter pylori  
US-09-895-913A-246

Query Match 37.3%; Score 38; DB 9; Length 1185;  
Best Local Similarity 43.8%; Pred. No. 2e+02;  
Matches 7; Conservative 4; Mismatches 3; Indels 2; Gaps 1;

QY 2 TSZPC--GAVMILYFH 15  
1::11 1 :11:1  
Db 1137 TAKPCLNGRFELLYH 1152

## RESULT 7

US-09-815-242-11466  
; Sequence 11466, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlssen, Karl L.  
; APPLICANT: zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; TITLE OF INVENTION: Prokaryotes  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 11466  
; LENGTH: 1185  
; TYPE: PRT  
; ORGANISM: Helicobacter pylori  
US-09-815-242-11466

Query Match 37.3%; Score 38; DB 10; Length 1185;  
Best Local Similarity 43.8%; Pred. No. 2e+02;  
Matches 7; Conservative 4; Mismatches 3; Indels 2; Gaps 1;

QY 2 TSZPC--GAVMILYFH 15  
1::11 1 :11:1  
Db 1137 TAKPCLNGRFELLYH 1152

## RESULT 8

US-09-864-761-37111  
; Sequence 37111, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO  
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
; FILE REFERENCE: Aeomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263,6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1

```
; SEQ ID NO 37111
; LENGTH: 111
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL023281.1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 9.2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 12
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 9.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 10
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 9.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 12
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 11
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 10
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 16
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 13
; OTHER INFORMATION: EST_HUMAN HIT: AW411290.1, EVALUE 5.00e-05
; OTHER INFORMATION: SWISSPROT HIT: P81274, EVALUE 9.00e-06
US-09-864-761-37111

Query Match      36.3%; Score 37; DB 10; Length 111;
Best Local Similarity 42.9%; Pred. No. 31;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 1 BTSZPCGAVMILYF 14
: : : 1 1 1 1 1 1
Db 85 DVAKTLGRVMLYF 98

RESULT 9
US-10-028-051-8
; Sequence 8, Application US/10028051
; Patent No. US20020147329A1
; GENERAL INFORMATION:
; APPLICANT: Luyten, Frank P.
; Hoang, Bang
; Moos, Jr., Malcolm
; Wang, Shouwen
; TITLE OF INVENTION: ISOLATION AND USE OF TISSUE
; GROWTH INDUCING FRZB PROTEIN
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive, 16th Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/028,051
; FILING DATE: 19-Dec-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/729,452
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Bartfeld, Neil S
; REGISTRATION NUMBER: 39,901
; REFERENCE/DOCKET NUMBER: NIH133.001A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-235-8550
; TELEFAX: 619-235-0176
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 318 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
```

```
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-10-028-051-8

Query Match      36.3%; Score 37; DB 12; Length 318;
Best Local Similarity 35.7%; Pred. No. 84;
Matches 5; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

OY 4 ZPCGAVMILYFHKW 17
: : : 1 1 1 1 1
Db 114 QGCEPILIKYRHSW 127

RESULT 10
US-09-903-170C-7
; Sequence 7, Application US/09903170C
; Patent No. US20020156249A1
; GENERAL INFORMATION:
; APPLICANT: De Robertis, Edward M.
; Boumeester, Tewls
; TITLE OF INVENTION: Endoderm, Cardiac and Neural Inducing
; FILE REFERENCE: 510015-259
; CURRENT APPLICATION NUMBER: US/09/903,170C
; CURRENT FILING DATE: 2001-07-11
; PRIOR APPLICATION NUMBER: US 60/020,150
; PRIOR FILING DATE: 1996-06-20
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Mouse FRZB-1
US-09-903-170C-7

Query Match      36.3%; Score 37; DB 9; Length 323;
Best Local Similarity 35.7%; Pred. No. 85;
Matches 5; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

OY 4 ZPCGAVMILYFHKW 17
: : : 1 1 1 1 1
Db 117 QGCEPILIKYRHSW 130

RESULT 11
US-09-903-180B-7
; Sequence 7, Application US/09903180B
; Patent No. US20020099171A1
; GENERAL INFORMATION:
; APPLICANT: De Robertis, Edward M.
; Boumeester, Tewls
; TITLE OF INVENTION: Endoderm, Cardiac and Neural Inducing
; FILE REFERENCE: 510015-256
; CURRENT APPLICATION NUMBER: US/09/903,180B
; CURRENT FILING DATE: 2001-07-11
; PRIOR APPLICATION NUMBER: US 60/020,150
; PRIOR FILING DATE: 1996-06-20
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Mouse FRZB-1
US-09-903-180B-7

Query Match      36.3%; Score 37; DB 10; Length 323;
Best Local Similarity 35.7%; Pred. No. 85;
Matches 5; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

OY 4 ZPCGAVMILYFHKW 17
: : : 1 1 1 1 1
Db 117 QGCEPILIKYRHSW 130
```

```
RESULT 12
US-09-903-187A-7
; Sequence 7, Application US/09903187A
; Patent No. US20020099172A1
; GENERAL INFORMATION:
; APPLICANT: De Robertis, Edward M.
; APPLICANT: Boumeester, Lewis
; TITLE OF INVENTION: Endoderm, Cardiac and Neural Inducing
; TITLE OF INVENTION: Factors
; FILE REFERENCE: 510015-248
; CURRENT APPLICATION NUMBER: US/09/903,187A
; CURRENT FILING DATE: 2001-07-11
; PRIOR APPLICATION NUMBER: 09/552,988
; PRIOR FILING DATE: 2000-04-21
; PRIOR APPLICATION NUMBER: US 08/878,474
; PRIOR FILING DATE: 1997-06-18
; PRIOR APPLICATION NUMBER: US 60/020,150
; PRIOR FILING DATE: 1996-06-20
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Mouse
US-09-903-187A-7

Query Match      36.3%; Score 37; DB 10; Length 323;
Best Local Similarity 35.7%; Pred. No. 85;
Matches 5; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY      4 ZPCGAVMILYFHKW 17
; :|:| | | |
Db      117 QGCEPILIKYRHSW 130

RESULT 13
US-09-903-171A-7
; Sequence 7, Application US/09903171A
; Patent No. US20020123613A1
; GENERAL INFORMATION:
; APPLICANT: De Robertis, Edward M.
; APPLICANT: Boumeester, Lewis
; TITLE OF INVENTION: Endoderm, Cardiac and Neural Inducing
; TITLE OF INVENTION: Factors
; FILE REFERENCE: 510015-260
; CURRENT APPLICATION NUMBER: US/09/903,171A
; CURRENT FILING DATE: 2001-07-11
; PRIOR APPLICATION NUMBER: US 60/020,150
; PRIOR FILING DATE: 1996-06-20
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Mouse FRZB-1
US-09-903-171A-7

Query Match      36.3%; Score 37; DB 10; Length 323;
Best Local Similarity 35.7%; Pred. No. 85;
Matches 5; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY      4 ZPCGAVMILYFHKW 17
; :|:| | | |
Db      117 QGCEPILIKYRHSW 130

RESULT 14
US-09-903-188A-7
; Sequence 7, Application US/09903188A
; Patent No. US20020128439A1
; GENERAL INFORMATION:
```

```
; APPLICANT: De Robertis, Edward M.
; APPLICANT: Boumeester, Lewis
; TITLE OF INVENTION: Endoderm, Cardiac and Neural Inducing
; TITLE OF INVENTION: Factors
; FILE REFERENCE: 510015-258
; CURRENT APPLICATION NUMBER: US/09/903,188A
; CURRENT FILING DATE: 2001-07-11
; PRIOR APPLICATION NUMBER: US 60/020,150
; PRIOR FILING DATE: 1996-06-20
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Mouse FRZB-1
US-09-903-188A-7

Query Match      36.3%; Score 37; DB 10; Length 323;
Best Local Similarity 35.7%; Pred. No. 85;
Matches 5; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY      4 ZPCGAVMILYFHKW 17
; :|:| | | |
Db      117 QGCEPILIKYRHSW 130

RESULT 15
US-09-903-323A-7
; Sequence 7, Application US/09903323A
; Patent No. US20020128440A1
; GENERAL INFORMATION:
; APPLICANT: De Robertis, Edward M.
; APPLICANT: Boumeester, Lewis
; TITLE OF INVENTION: Endoderm, Cardiac and Neural Inducing
; TITLE OF INVENTION: Factors
; FILE REFERENCE: 510015-261
; CURRENT APPLICATION NUMBER: US/09/903,323A
; CURRENT FILING DATE: 2001-07-11
; PRIOR APPLICATION NUMBER: US 60/020,150
; PRIOR FILING DATE: 1996-06-20
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Mouse FRZB-1
US-09-903-323A-7

Query Match      36.3%; Score 37; DB 10; Length 323;
Best Local Similarity 35.7%; Pred. No. 85;
Matches 5; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY      4 ZPCGAVMILYFHKW 17
; :|:| | | |
Db      117 QGCEPILIKYRHSW 130
```

Search completed: November 15, 2002, 11:48:26  
Job time : 11 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: November 15, 2002, 11:43:14 ; Search time 15 Seconds  
(without alignments)  
115.361 Million cell updates/sec

Title: US-09-881-569A-2

Perfect score: 102

Sequence: 1 BRSZPCGAVMILYFHKWR 18

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Minimum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_73:\*  
1: pirl:\*  
2: pirl2:\*  
3: pirl3:\*  
4: pirl4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description                |
|------------|-------|-------------|--------|-------|----------------------------|
| 1          | 44    | 43.1        | 222    | 2     | B69433 signal-transducing  |
| 2          | 44    | 43.1        | 758    | 2     | A47439 gamma-glutamyl car  |
| 3          | 43    | 42.2        | 64     | 2     | AC2277 hypothetical prote  |
| 4          | 43    | 42.2        | 1239   | 2     | G71266 probable ATP-depen  |
| 5          | 42    | 41.2        | 221    | 2     | D86600 SET domain protein  |
| 6          | 42    | 41.2        | 221    | 2     | B81514 conserved hypotet   |
| 7          | 42    | 41.2        | 221    | 2     | D72024 set domain protein  |
| 8          | 42    | 41.2        | 302    | 2     | AE0958 conserved lysr-fam1 |
| 9          | 42    | 41.2        | 443    | 2     | H83589 conserved hypotet   |
| 10         | 41    | 40.2        | 353    | 2     | D84686 probable RAD51B-11  |
| 11         | 41    | 40.2        | 374    | 2     | T09771 phosphatidate cyti  |
| 12         | 41    | 40.2        | 544    | 2     | T05952 Mio-h1 protein - b  |
| 13         | 41    | 40.2        | 558    | 2     | T30418 chitinase (EC 3.2.  |
| 14         | 41    | 40.2        | 575    | 2     | AB1793 ABC transporter (A  |
| 15         | 41    | 40.2        | 575    | 2     | AH1417 ABC transporter (A  |
| 16         | 41    | 40.2        | 612    | 2     | F83000 ABC transporter (A  |
| 17         | 41    | 40.2        | 623    | 2     | D88922 probable two-compo  |
| 18         | 40.5  | 39.7        | 2133   | 2     | T30637 protein W03F9.4 [1  |
| 19         | 40    | 39.2        | 60     | 1     | TIBOR hypothetical prote   |
| 20         | 40    | 39.2        | 100    | 2     | S45888 serum basic protel  |
| 21         | 40    | 39.2        | 195    | 2     | AE3074 hypothetical prote  |
| 22         | 40    | 39.2        | 199    | 2     | D98212 transcription regu  |
| 23         | 40    | 39.2        | 272    | 2     | I64245 probable tetr-fami  |
| 24         | 40    | 39.2        | 536    | 2     | T33480 hypothetical prote  |
| 25         | 40    | 39.2        | 590    | 2     | A10094 ABC-transporter tr  |
| 26         | 40    | 39.2        | 830    | 1     | A57060 ABC-transporter tr  |
| 27         | 39.5  | 38.7        | 340    | 2     | AE0353 serine/threonine-s  |
| 28         | 39.5  | 38.7        | 737    | 1     | JC5695 probable membrane   |
| 29         | 39.5  | 38.7        | 1010   | 2     | T33372 Dnm1p/Vps1p-like p  |
|            |       |             |        |       | hypothetical prote         |

|    |      |      |      |   |                           |
|----|------|------|------|---|---------------------------|
| 30 | 39.5 | 38.7 | 1157 | 2 | AD1358 ATP-dependent deox |
| 31 | 39.5 | 38.7 | 1157 | 2 | AD1728 ATP-dependent deox |
| 32 | 39   | 38.2 | 148  | 2 | D69541 flavodoxin homolog |
| 33 | 39   | 38.2 | 284  | 2 | H86898 undecaprenol kinas |
| 34 | 39   | 38.2 | 328  | 2 | F89914 hypothetical prote |
| 35 | 39   | 38.2 | 388  | 1 | WMLJBT bel-2 protein - sl |
| 36 | 39   | 38.2 | 422  | 2 | F70018 multiple sugar-bin |
| 37 | 39   | 38.2 | 486  | 2 | T24334 hypothetical prote |
| 38 | 39   | 38.2 | 500  | 1 | JE0259 cytochrome P450 2D |
| 39 | 39   | 38.2 | 625  | 2 | T41603 alpha-amylase - fl |
| 40 | 39   | 38.2 | 738  | 2 | C95936 conserved hypotet  |
| 41 | 39   | 38.2 | 860  | 2 | C72338 hypothetical prote |
| 42 | 39   | 38.2 | 1147 | 2 | S64930 serine/threonine-s |
| 43 | 38.5 | 37.7 | 75   | 2 | A31298 alpha-amylase inh  |
| 44 | 38.5 | 37.7 | 227  | 2 | I37019 cytochrome-c oxid  |
| 45 | 38.5 | 37.7 | 255  | 2 | AF1556 conserved hypotet  |

## ALIGNMENTS

## RESULT 1

B69433

signal-transducing histidine kinase homolog - Archaeoglobus fulgidus

C/Species: Archaeoglobus fulgidus

C/Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 21-Jul-2000

C/Accession: B69433

R/Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod

-; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E

Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.

Nature 390, 364-370, 1997

A/Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artlach, P.; Kaine, B.P.; Sykes,

Smith, H.O.; Woese, C.R.; Venter, J.C.

A/Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch

A/Reference number: A69250; MUID:98049343; PMID:9389475

A/Accession: B69433

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-222 &lt;KLE&gt;

A/Cross-references: GB:AE001001; GB:AE000782; NID:g2689324; PIDN:AAB89779.1; PID:g264

## Query Match

Best Local Similarity 43.1%; Score 44; DB 2; Length 222;

Matches 6; Conservativity 50.0%; Pred. No. 7.4;

Matches 6; Conservativity 4; Mismatches 2; Indels 0; Gaps 0;

QY 3 SZPCGAVMILYF 14

DB 204 NEPCGAVFVMEF 215

## RESULT 2

A47439

gamma-glutamyl carboxylase (EC 4.1.1.-) - bovine

C/Species: Bos primigenius taurus (cattle)

C/Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 05-Nov-1999

C/Accession: A47439; B39283; C39283

R/Rehmetulla, A.; Roth, D.A.; Wasley, L.C.; Kuliopulos, A.; Walsh, C.T.; Furie, B.; F

Proc. Natl. Acad. Sci. U.S.A. 90, 4611-4615, 1993

A/Title: In vitro and in vivo functional characterization of bovine vitamin K-depende

A/Reference number: A47439; MUID:93281629; PMID:8506307

A/Accession: A47439

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-758 &lt;REH&gt;

A/Cross-references: GB:L09726; NID:g289398; PIDN:AAA30425.1; PID:g289399

R/Wu, S.M.; Cheung, W.F.; Frazier, D.; Stafford, D.W.

submitted to GenBank, January 1992

A/Reference number: A38939

A/Accession: A38939

A/Molecule type: mRNA

A/Residues: 96-758 &lt;WUA&gt;

A/Cross-references: GB:M81593; NID:g162744; PIDN:AAA30410.1; PID:g162745

R/Wu, S.M.; Cheung, W.F.; Frazier, D.; Stafford, D.W.

Science 254, 1634-1636, 1991

A:Title: Cloning and expression of the cDNA for human gamma-glutamyl carboxylase.

A:Reference number: A39283; MUID:92086858; PMID:1749935

A:Accession: B39283

A>Status: nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 96-693,695-758 <WU2>

A:Cross-references: GB:M81593

A:Accession: C39283

A:Molecule type: protein

A:Residues: 37-50 <WU3>

C:Keywords: carbon-carbon lyase; carboxy-lyase

Query Match 43.1%; Score 44; DB 2; Length 758;

Best Local Similarity 46.7%; Pred. No. 22;

Matches 7; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 BPSZPCGAVMILYFH 15

DB 273 DASRPICGFVFVSYFH 287

### RESULT 3

AC2277

hypothetical protein asr3770 [imported] - Nostoc sp. (strain PCC 7120)

C:Species: Nostoc sp.

A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C>Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002

C:Accession: AC2277

R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kurlitz, T.; Sasamoto, S.; Watanabe, A.; Triguero, N.; Shimizu, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.

DNA Res. 8, 205-213, 2001

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena

A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: AC2277

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-64 <KUR>

A:Cross-references: GB:BA000019; PIDN:BA875469.1; PID:g17132904; GSPDB:GN00179

A:Experimental source: strain PCC 7120

C:Genetics:

A:Gene: asr3770

Query Match 42.2%; Score 43; DB 2; Length 64;

Best Local Similarity 58.3%; Pred. No. 3.5;

Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 6 CGAVMILYFHKW 17

DB 12 CCGVREKYEYFHW 23

### RESULT 4

G71266

probable ATP-dependent nuclease, subunit A - syphilis spirochete

C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)

C>Date: 24-Jul-1998 #sequence\_revision 24-Jul-1998 #text\_change 05-Nov-1999

C:Accession: G71266

R;Fraser, C.M.; Norris, S.J.; Weinstein, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin, J.; Khatkhat, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McDevitt, L.; Weidman, J.; Smith, H.O.; Venter, J.C.

Science 281, 375-388, 1998

A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.

A:Reference number: A71250; MUID:98332770; PMID:9665876

A:Accession: G71266

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-1239 <COL>

A:Cross-references: GB:AE001259; GB:AE000520; MID:g3323209; PIDN:AAC65853.1; PID:g332321

A:Experimental source: strain Nichols

C:Genetics:

A:Gene: TP0898

Query Match 42.2%; Score 43; DB 2; Length 1239;

Best Local Similarity 53.3%; Pred. No. 52;

Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 2 TSZPCGAVMILYFHK 16

DB 833 TDRACGAVQIMSVHK 847

### RESULT 5

D86600

SET domain protein [imported] - Chlamydia pneumoniae (strain J138)

C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae

C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 02-Mar-2001

C:Accession: D86600

R;Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.

Nucleic Acids Res. 28, 2311-2314, 2000

A:Title: Comparison of whole genome sequences of Chlamydia pneumoniae J138.

A:Reference number: A86491; MUID:20330349; PMID:10871362

A:Accession: D86600

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-221 <STO>

A:Cross-references: GB:BA000008; MID:g8979252; PIDN:BAA99086.1; GSPDB:GN00142

A:Experimental source: strain J138

C:Genetics:

A:Gene: CPJ0878

Query Match 41.2%; Score 42; DB 2; Length 221;

Best Local Similarity 35.3%; Pred. No. 16;

Matches 6; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 2 TSZPCGAVMILYFHKW 18

DB 5 TTEPCSSIHSLNNDWR 21

### RESULT 6

B81514

conserved hypothetical protein CP0991 [imported] - Chlamydia pneumoniae (strain A

C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae

C>Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 11-May-2000

C:Accession: B81514

R;Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke, C.; Dodson, R.; Gwin, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg, S.

Nucleic Acids Res. 28, 1397-1406, 2000

A:Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39

A:Reference number: A81500; MUID:20150255; PMID:10684935

A:Accession: B81514

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-221 <REA>

A:Cross-references: GB:AE002257; GB:AE002161; MID:g7189902; PIDN:AAF38770.1; PID:g718

A:Experimental source: strain AR39, HL cells

C:Genetics:

A:Gene: CP0991

Query Match 41.2%; Score 42; DB 2; Length 221;

Best Local Similarity 35.3%; Pred. No. 16;

Matches 6; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 2 TSZPCGAVMILYFHKW 18

DB 5 TTEPCSSIHSLNNDWR 21

### RESULT 7

D72024

set domain protein - Chlamydia pneumoniae (strain CWL029)

C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae

C>Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 05-May-2000

C:Accession: D72024

R;Kalman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood,





C:Keywords: nucleotidyltransferase

Query Match 40.2%; Score 41; DB 2; Length 374;  
Best Local Similarity 45.5%; Pred. No. 38;  
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 7 GAVMILYFHKW 17  
|::|::|  
Db 103 GTEVVVYFYK 113

## RESULT 12

T05952

Mlo-hl protein - barley

C:Species: Hordeum vulgare (barley)

C:Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 15-Jun-2001

C:Accession: T05952

;Panstruga, R.

submitted to the EMBL Data Library, May 1997

A:Reference number: 215093

A:Accession: T05952

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-544 &lt;PAN&gt;

A:Cross-references: EMBL:Z95496; PIDN:CAB08860.1

A:Experimental source: cv. Igri

C:Genetics:

A:Gene: Mlo-hl

A:Map position: 4H

A:Introns: 41/3; 139/3; 178/3; 199/1; 282/2; 299/1; 312/3; 382/3

C:Superfamily: barley pathogen resistance protein Mlo

Query Match 40.2%; Score 41; DB 2; Length 544;  
Best Local Similarity 40.0%; Pred. No. 53;  
Matches 10; Conservative 1; Mismatches 0; Indels 14; Gaps 1;

OY 8 AVMIL-----YFHKW 18  
|::|::|  
Db 23 AVMILSVAMEHALHKLGHFHKW 47

## RESULT 13

T30418

chitinase (EC 3.2.1.14) - Lymantria dispar nuclear polyhedrosis virus

C:Species: Lymantria dispar nuclear polyhedrosis virus, LdMNPV

C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 21-Jul-2000

C:Accession: T30418

R;Kuzio, J.; Pearson, M.N.; Harwood, S.H.; Funk, C.J.; Evans, J.T.; Slavecek, J.M.; Rohr

Virology 253, 17-34, 1999

A:Title: Sequence and analysis of the genome of a baculovirus pathogenic for Lymantria d

A:Reference number: Z20836; MUID:99124785; PMID:9887315

A:Accession: T30418

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-558 &lt;KU2&gt;

A:Cross-references: EMBL:AF081810; NID:g3822234; PIDN:AAC70256.1; PID:g3822305

C:Function:

A:Description: catalyzes the hydrolysis of chitin, a beta-1,4-linked homopolymer of N-ac

A:Pathway: chitin degradation

C:Superfamily: serrata marcescens chitinase

C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 40.2%; Score 41; DB 2; Length 558;  
Best Local Similarity 40.0%; Pred. No. 54;  
Matches 6; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

OY 3 SZPGAVMILYFHKW 17  
::|::|::|  
Db 145 AEPTCKVVAAYFEW 159

## RESULT 14

ABI793

ABC transporter (ATP-binding protein) homolog lln2888 [imported] - Listeria innocua (

C:Species: Listeria innocua

C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001

C:Accession: ABI793

R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec

.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsthl,

D.; Jones, L.M.; Karst, U.

Science 294, 849-852, 2001

A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkut, G.; Madueno, E.; Maltournam, A.;

ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla

A:Title: Comparative genomics of Listeria species.

A:Reference number: ABI077; MUID:21537279; PMID:11679669

A:Accession: ABI793

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-575 &lt;GLA&gt;

A:Cross-references: GB:AL592022; PIDN:CAC98114.1; PID:gl6415423; GSPDB:GN00178

A:Experimental source: strain Clp11262

C:Genetics:

A:Gene: lln2888

C:Superfamily: Escherichia coli ABC transporter mdla; ATP-binding cassette homology

Query Match 40.2%; Score 41; DB 2; Length 575;  
Best Local Similarity 50.0%; Pred. No. 56;  
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 7 GAVMILYFHKW 18  
|::|::|  
Db 150 GATILFFMDWK 161

## RESULT 15

AH1417

ABC transporter (ATP-binding protein) homolog lmo2745 [imported] - Listeria monocytog

C:Species: Listeria monocytogenes

C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001

C:Accession: AH1417

R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec

.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsthl,

D.; Jones, L.M.; Karst, U.

Science 294, 849-852, 2001

A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkut, G.; Madueno, E.; Maltournam, A.;

ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla

A:Title: Comparative genomics of Listeria species.

A:Reference number: ABI077; MUID:21537279; PMID:11679669

A:Accession: AH1417

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-575 &lt;GLA&gt;

A:Cross-references: GB:NC\_003210; PIDN:CAD00958.1; PID:gl6412245; GSPDB:GN00177

A:Experimental source: strain EGD-e

C:Genetics:

A:Gene: lmo2745

C:Superfamily: Escherichia coli ABC transporter mdla; ATP-binding cassette homology

Query Match 40.2%; Score 41; DB 2; Length 575;  
Best Local Similarity 50.0%; Pred. No. 56;  
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 7 GAVMILYFHKW 18  
|::|::|  
Db 150 GATILFFMDWK 161

Search completed: November 15, 2002, 11:44:35  
Job time : 17 secs





DB 273 DASRPICLVFVSFHH 287

RESULT 2

VKGC\_BOVIN

ID VKGC\_BOVIN STANDARD: PRT: 758 AA.

AC 007175;

DT 01-OCT-1994 (Rel. 30, Created)

DT 01-OCT-1994 (Rel. 30, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE Vitamin K-dependent gamma-carboxylase (EC 6.4.-.-) (Gamma-glutamyl carboxylase).

GN GGCX OR GC.

OS Bos taurus (Bovine).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos.

NCBI\_TaxID=9913;

RN [1]

RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

RC TISSUE=Liver;

RX MEDLINE=93281629; PubMed=8506307;

RA Rehenculla A., Roth D.A., Wasley L.C., Kullipulos A., Walsh C.T., Furtle B., Furtle B.C., Kaufman R.J.;

RT "In vitro and in vivo functional characterization of bovine vitamin K-dependent gamma-carboxylase expressed in Chinese hamster ovary cells."

RL Proc. Natl. Acad. Sci. U.S.A. 90:4611-4615(1993).

RN [2]

RP SEQUENCE OF 96-758 FROM N.A., AND PARTIAL SEQUENCE.

RC TISSUE=Liver;

RX MEDLINE=92086858; PubMed=1749935;

RA Wu S.-M., Chung W.-F., Frazier D., Stafford D.W.;

RT "Cloning and expression of the cDNA for human gamma-glutamyl carboxylase."

RL Science 254:1634-1636(1991).

CC -1- FUNCTION: VITAMIN K-DEPENDENT CARBOXYLATION OF MULTIPLE AMINO-TERMINAL GLUTAMIC ACID RESIDUES IN VARIOUS PROTEINS, IT CONVERTS GLUTAMATE RESIDUES TO GAMMA-CARBOXYGLUTAMATE.

CC -1- SUBUNIT: MAY FORM A HOMODIMER (PROBABLE).

CC -1- PTM: THE N-TERMINUS IS BLOCKED.

CC -1- SIMILARITY: SOME, TO PLANT AND MAMMALIAN LIPOXYGENASES.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL: L09726; AAA30425.1; -

DR EMBL: M81593; AAA30410.1; -

DR InterPro: IPR001870; Gamma\_carboxylase.

KW Ligase.

SO SEQUENCE 758 AA; 87557 MW; 9B7F47B809D1A873 CRC64;

Query Match 43.1%; Score 44; DB 1; Length 758;

Best Local Similarity 46.7%; Pred. NO. 9.6;

Matches 7; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

OY 1 BRSZPCGAVMILYFH 15

DB 273 DASRPICLVFVSFHH 287

RESULT 3

CDSA\_MYCGE STANDARD: PRT: 305 AA.

ID CDSA\_MYCGE

AC 049433; 049309;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Putative phosphatidate cytidyltransferase (EC 2.7.7.41) (CDP-diglyceride synthetase) (CDP-diglyceride pyrophosphorylase) (CDP-diacylglycerol synthase) (CDS) (CTP:phosphatidate cytidyltransferase) (CDP-DAG synthase) (CDP-DG synthetase).

GN CDSA OR MG437.

OS Mycoplasma genitalium.

OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.

NCBI\_TaxID=2097;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 33530 / G-37;

RX MEDLINE=96026346; PubMed=7569993;

RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A., Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M., Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L., Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M., Tomb J.-F., Dougherty B.A., Bort K.F., Hu P.-C., Lucier T.S., Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;

RT "The minimal gene complement of Mycoplasma genitalium."

RL Science 270:397-403(1995).

RN [2]

RP SEQUENCE OF 127-226 FROM N.A.

RC STRAIN=ATCC 33530 / G-37;

RX MEDLINE=94075230; PubMed=8253680;

RA Peterson S.N., Hu P.-C., Bort K.F., Hutchison C.A. III;

RT "A survey of the Mycoplasma genitalium genome by using random sequencing."

RL J. Bacteriol. 175:7918-7930(1993).

CC -1- CATALYTIC ACTIVITY: CTP + phosphatidate -> diphosphate + CDP-diacylglycerol.

CC -1- PATHWAY: Phospholipid biosynthesis.

CC -1- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).

CC -1- SIMILARITY: BELONGS TO THE CDS FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL: U39725; AAC72458.1; ALT\_INIT.

DR EMBL: U02189; AAD12473.1; -

DR TIGR: MG437; -

DR InterPro: IPR000374; Pcttransf.

DR Pfam: PF01148; Cytidylyltrans; 1.

DR PROSITE: PS01315; CDS; 1.

KW Hypothetical protein; Transferase; Nucleotidyltransferase; Phospholipid biosynthesis; Transmembrane; Complete proteome.

KW

FT TRANSMEM 22 42 POTENTIAL.

FT TRANSMEM 51 71 POTENTIAL.

FT TRANSMEM 84 104 POTENTIAL.

FT TRANSMEM 107 127 POTENTIAL.

FT TRANSMEM 152 172 POTENTIAL.

FT TRANSMEM 212 232 POTENTIAL.

FT TRANSMEM 264 284 POTENTIAL.

SO SEQUENCE 305 AA; 34195 MW; F149FA6BD9DDC080 CRC64;

Query Match 40.2%; Score 41; DB 1; Length 305;

Best Local Similarity 45.5%; Pred. NO. 13;

Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 7 GAVMILYFHKM 17

DB 34 GTFVVVYFYKM 44

RESULT 4

MLH1\_HORVU STANDARD: PRT: 544 AA.

ID MLH1\_HORVU

AC 049873;

DT 15-JUL-1999 (Rel. 38, Created)

DT 15-JUL-1999 (Rel. 38, last sequence update)  
 DT 15-JUL-1999 (Rel. 38, last annotation update)  
 DE MLO protein homolog 1.  
 GN MLO-H1.  
 OS Hordeum vulgare (Barley).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;  
 OC Triticeae; Hordeum.  
 OX NCBI\_TaxID=4513;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Igri;  
 RA Panstruga R.;  
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1 SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
 CC -1 SIMILARITY: BELONGS TO THE MLO FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

| DR | InterPro: IPR004326; MIO.                          |            |
|----|--|------------|
| DR | Pfam: PF03094; MIO; 1.                             |            |
| KW | Transmembrane.                                     |            |
| FT | TRANSMEM 16  | POTENTIAL. |
| FT | TRANSMEM 285                                       | POTENTIAL. |
| FT | TRANSMEM 403                                       | POTENTIAL. |
| SO | SEQUENCE 544 AA; 60875 MW; D4FBCDD2A2AB7796 CRC64; |            |

|                          |        |               |            |             |
|--------------------------|--------|---------------|------------|-------------|
| Query Match              | 40.28; | Score 41;     | DB 1;      | Length 544; |
| Best Local Similarity    | 40.08; | Pred. No. 22; |            |             |
| Matches 10; Conservative | 1;     | Mismatches 0; | Indels 14; | Gaps 1;     |

```

      Oy      8  AMVIL-----YFHKWR 18
                |||||
      Db      23  AMVILSVAMENALHKLGHWFHKWR 47

```

|            |   |                                   |
|------------|---|-----------------------------------|
| RESULT 5   |   |                                   |
| IBPS_BOVIN |   |                                   |
| ID         | IBPS_BOVIN                                  | STANDARD;                         |
| AC         | P00975;                                     | PRT; 60 AA.                       |
| DT         | 21-JUL-1986                                 | (Rel. 01, Created)                |
| DT         | 21-JUL-1986                                 | (Rel. 01, Last sequence update)   |
| DT         | 16-OCT-2001                                 | (Rel. 40, Last annotation update) |
| DE         | Serum basic protease inhibitor (Serum BPI). |                                   |

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovidae; Bovinae; Bos.  
NCBI\_TaxID=9913;  
OX

RP SEQUENCE.  
RX MEDLINE-81044408; PubMed-7428928;  
RA Wachter E., Depner K., Hochstrasser K., Lempart K., Geiger R.;  
RT "A new Kunitz-type inhibitor from bovine serum amino acid sequence  
RL determination.";  
RL FEBS Lett. 119:58-62(1980).  
CC -!- FUNCTION: THIS INHIBITOR HAS ACTIVITY VERY SIMILAR TO THAT OF THE  
BASIC PROTEASE INHIBITOR FROM BOVINE TISSUES.

```
CC      -1- SIMILARITY: CONTAINS 1 BPTI/KUNITZ INHIBITOR DOMAIN
DR      PIR; A01206; TBOR.
DR      HSSP; P00974; IBPI.
DR      InterPro; IPR002223; Kunitz_BPTI.
DR      Pfam; PF00014; Kunitz_BPTI; 1.
DR      PRINTS; PR00759; BASICPTASE.
DR      ProDom; PD000222; Kunitz_BPTI; 1.
```

| Accession                              | Protein Name                        | Length (aa) | MW (kDa) | CD                      |
|--|-------------------------------------|-------------|----------|-------------------------|
| DR SMART; SM00131; KU; 1.              | SMART; SM00131; KU; 1.              | 60          | 6.647    | B9953EBAACF1A4E6 CRC64; |
| DR PROSITE; PS00280; BPTI_KUNITZ_1; 1. | PROSITE; PS00280; BPTI_KUNITZ_1; 1. | 57          | 6.647    | B9953EBAACF1A4E6 CRC64; |
| DR PROSITE; PS00279; BPTI_KUNITZ_2; 1. | PROSITE; PS00279; BPTI_KUNITZ_2; 1. | 57          | 6.647    | B9953EBAACF1A4E6 CRC64; |
| KW Serine protease inhibitor.          | Serine protease inhibitor.          | 57          | 6.647    | B9953EBAACF1A4E6 CRC64; |
| FT DISULFID 7                          | DISULFID 7                          | 7           | 6.647    | B9953EBAACF1A4E6 CRC64; |
| FT DISULFID 16                         | DISULFID 16                         | 16          | 6.647    | B9953EBAACF1A4E6 CRC64; |
| FT DISULFID 32                         | DISULFID 32                         | 32          | 6.647    | B9953EBAACF1A4E6 CRC64; |
| FT ACT_SITE 17                         | ACT_SITE 17                         | 17          | 6.647    | B9953EBAACF1A4E6 CRC64; |
| SO SEQUENCE                            | SEQUENCE                            | 60          | 6.647    | B9953EBAACF1A4E6 CRC64; |

|                       |        |                |       |               |
|-----------------------|--------|----------------|-------|---------------|
| Query Match           | 39.28; | Score 40;      | DB 1; | Length 60;    |
| Best Local Similarity | 63.68; | Pred. No. 3.7; |       |               |
| Matches               | 7;     | Conservative   | 1;    | Mismatches 3; |
|                       |        |                |       | Indels 0;     |
|                       |        |                |       | Gaps 0;       |

```
QY      5 PCGAVMILYFH 15
          |||||:
Db     15 PCKAAMIRYFY 25
```

```

RESULT 6
YB02_YEAST
ID YB02_YEAST STANDARD; PRT; 100 AA.
AC P38323;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Hypothetical 11.5 kDa protein in RPL4A-HMT1 intergenic region.
GN YBR032W OR YBR0317.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;

```

RP SEQUENCE FROM N.A.  
RC STRAIN-S28BC;  
RX MEDLINE=94378725; Pubmed=8091864;  
RA Smits P.H.M., de Haan M., Maat C., Grivell L.A.;  
RT "The complete sequence of a 33 kb fragment on the right arm of  
RT chromosome II from *Saccharomyces cerevisiae* reveals 16 open reading  
RT frames, including ten new open reading frames, five previously  
RT identified genes and a homologue of the *SCO1* gene.";   
RL Yeast 10:S75-S80(1994).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL, Z35901; CAA84974.1; -.  
DR PIR, S45888; S45888.  
DR SCD, S0000236; YBR032W.  
KW Hypothetical protein.  
SQ SEQUENCE 100 AA; 11485 MW; 1F2F4DEAEC6C2793 CRC64;

|                       |        |                |       |               |
|-----------------------|--------|----------------|-------|---------------|
| Query Match           | 39.28; | Score 40;      | DB 1; | Length 100;   |
| Best Local Similarity | 40.08; | Pred. No. 6.2; |       |               |
| Matches               | 6;     | Conservative   | 5;    | Mismatches 4; |
|                       |        |                |       | Indels 0;     |
|                       |        |                |       | Gaps 0;       |

QY 1 BTSZPCGAVMILYFH 15  
:|:|:|:|:|  
Db 33 DTSQPIGSMQVLLPH 47

| RESULT 7 |             |                                   |
|----------|-------------|-----------------------------------|
| ID       | Y415_MYCGE  |                                   |
| AC       | P47655;     | STANDARD; PRT; 272 AA.            |
| DT       | 01-FEB-1996 | (Rel. 33, Created)                |
| DT       | 01-FEB-1996 | (Rel. 33, Last sequence update)   |
| DT       | 16-OCT-2001 | (Rel. 40, Last annotation update) |

DE Hypothetical protein MG415.  
GN MG415.  
OS Mycoplasma genitalium.  
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.  
OX NCBI\_TaxID=2097;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 33530 / G-37;  
RX MEDLINE=96026346; PubMed=7569993;  
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,  
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,  
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,  
RA Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,  
RA Tomb J.-F., Dougherty B.A., Bolt K.F., Hu P.-C., Lucier T.S.,  
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;  
RT "The minimal gene complement of Mycoplasma genitalium.";  
SC Science 270:397-403(1995).  
CC -1- SIMILARITY: BELONGS TO THE MG414 / MG415 FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC  
CC  
CC EMBL: U39723; AAC71642.1; -  
CC TIGR: MG415; -  
CC Hypothetical protein; Transmembrane; Complete proteome.  
CC TRANSMEM 233 POTENTIAL.  
CC SEQUENCE 272 AA; 32504 MW; A221BAA6185231A4 CRC64;  
SO  
Query Match 39.2%; Score 40; DB 1; Length 272;  
Best Local Similarity 62.5%; Pred. No. 17;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 10 MILYFHKW 17  
Db 250 VFYFHKW 257  
-----  
RESULT 8  
UD13\_RAT STANDARD; PRT; 531 AA.  
ID UD13\_RAT  
AC 064637;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE UDP-glucuronosyltransferase 1-3 precursor, microsomal (EC 2.4.1.17)  
DE (UDPGT) (UGT1\*3) (UGT1-03) (UGT1A3) (B3).  
GN UGT1.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE OF 1-286 FROM N.A.  
RC STRAIN=Wistar;  
RX MEDLINE=95332265; PubMed=7608130;  
RA Eml Y., Ikushiro S.I., Iyanagi T.;  
RT "Drug-responsive and tissue-specific alternative expression of  
RT multiple first exons in rat UDP-glucuronosyltransferase family 1  
RT (UGT1) gene complex.";  
RL J. Biochem. 117:392-399(1995).  
RN [2]  
RP SEQUENCE OF 287-531 FROM N.A.  
RC TISSUE=Liver;  
RX MEDLINE=90274676; PubMed=2112380;  
RA Sato H., Koiwai O., Tanabe K., Kashiwamata S.;  
RT "Isolation and sequencing of rat liver bilirubin UDP-  
RT glucuronosyltransferase cDNA: possible alternate splicing of a common  
RT primary transcript.";

RL Biochem. Biophys. Res. Commun. 169:260-264(1990).  
CC -1- FUNCTION: UDPGT IS OF MAJOR IMPORTANCE IN THE CONJUGATION AND  
CC SUBSEQUENT ELIMINATION OF POTENTIALLY TOXIC XENOBIOTICS AND  
CC ENDOGENOUS COMPOUNDS.  
CC -1- CATALYTIC ACTIVITY: UDP-glucuronate + acceptor = UDP + acceptor  
CC beta-D-glucuronoside.  
CC -1- SUBCELLULAR LOCATION: Microsomal.  
CC -1- ALTERNATIVE PRODUCTS: THE UGT1 GENE ENCODES FOR MANY DIFFERENT.  
CC UDPGT ISOZYMES WHICH HAVE A DIFFERENT N-TERMINAL DOMAIN AND A  
CC COMMON C-TERMINAL DOMAIN OF 245 RESIDUES.  
CC -1- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC  
CC  
CC EMBL: D38067; BAA07262.1; -  
CC DR EMBL: M34007; AAA42312.1; ALT\_TERM.  
CC DR InterPro: IPR002213; UDPGT.  
CC DR Pfam: PF00201; UDPGT; 1.  
CC DR PROSITE: PS00375; UDPGT; 1.  
CC KW Transferase; Glycosyltransferase; Glycoprotein; Transmembrane; Signal;  
CC Multigene family; Microsome; Alternative splicing.  
CC FT SIGNAL 1 25 POTENTIAL.  
CC FT CHAIN 26 531 UDP-GLUCURONOSYLTRANSFERASE 1-3.  
CC FT TRANSMEM 489 505 POTENTIAL.  
CC FT CARBOHYD 116 116 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC FT CARBOHYD 139 139 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC FT CARBOHYD 293 293 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC FT CARBOHYD 431 431 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC SO SEQUENCE 531 AA; 60140 MW; ECB0994C6A2C7CEP CRC64;  
Query Match 39.2%; Score 40; DB 1; Length 531;  
Best Local Similarity 66.7%; Pred. No. 32;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 5 PCGAVMILY 13  
Db 153 PCGAVLAMY 161  
-----  
RESULT 9  
PKN2\_MYXXA STANDARD; PRT; 830 AA.  
ID PKN2\_MYXXA  
AC P54736;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Serine/threonine-protein kinase pkn2 (EC 2.7.1.-).  
DE PKN2.  
GN Myxococcus xanthus.  
OC Bacteria; Proteobacteria; delta subdivision; Myxobacteria;  
OC Myxococcales; Cystobacteriineae; Myxococcaceae; Myxococcus.  
OX NCBI\_TaxID=34;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=DZF1;  
RX MEDLINE=95293229; PubMed=7774814;  
RA Udo H., Munoz-Dorado J., Inouye M., Inouye S.;  
RT "Myxococcus xanthus, a Gram-negative bacterium, contains a  
RT transmembrane protein serine/threonine kinase that blocks the  
RT secretion of beta-lactamase by phosphorylation.";  
RL Genes Dev. 9:972-983(1995).  
RN [2]  
RP -1- FUNCTION: REGULATES THE ACTIVITY OF ENDOGENOUS BETA-LACTAMASE OR  
RP RELATED ENZYMES, BY BLOCKING THEIR SECRETION BY PHOSPHORYLATION.  
CC IN RESPONSE TO AN EXTERNAL SIGNAL YET TO BE IDENTIFIED.  
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
CC -1- SIMILARITY: CONTAINS 1 GUANYLATE CYCLASE DOMAIN.  
-----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M94857; AAA98813.1; -.  
DR Phosphatase; P54736; -.  
DR InterPro; IPR000719; Euk\_pkinase.  
DR InterPro; IPR001054; G\_cyclase.  
DR InterPro; IPR002290; Ser\_thr\_pkinase.  
DR Pfam; PF00069; pkinase; 1.  
DR Pfam; PF00211; guanylate\_cyc; 1.  
DR ProDom; PD000001; Euk\_pkinase; 1.  
DR SMART; SM00044; CYCC; 1.  
DR PROSITE; PS50125; GUANYLATE\_CYCLASES\_2; 1.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
KW Transferase; Serine/threonine-protein kinase; ATP-binding;  
KW Transmembrane.  
KW DOMAIN 1 605 CYTOPLASMIC (POTENTIAL).  
TRANSMEM 606 623 POTENTIAL.  
DOMAIN 624 830 PERIPLASMIC (POTENTIAL).  
FT DOMAIN 13 283 PROTEIN KINASE.  
FT DOMAIN 396 511 GUANYLATE CYCLASE.  
FT NP\_BIND 19 27 ATP (BY SIMILARITY).  
FT BINDING 42 42 ATP (BY SIMILARITY).  
FT ACT\_SITE 135 135 BY SIMILARITY.  
SQ SEQUENCE 830 AA; 87679 MW; BD53F33CE057FA1F CRC64;  
  
Query Match 39.2%; Score 40; DB 1; Length 830;  
Best Local Similarity 50.0%; Pred. No. 50;  
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;  
  
QY 2 TSZPCGAVMI 11  
I:::||||:  
Db 535 TAEPCGATTL 544  
  
RESULT 10  
IRX6\_MOUSE STANDARD; PRT; 438 AA.  
ID IRX6\_MOUSE  
AC Q9ER75;  
DT 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Iroquois-class homeodomain protein IRX-6 (Iroquois homeobox protein  
DE 6).  
GN IRX6.  
Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.  
OX NCBI\_Taxid=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NIH Swiss; TISSUE=Heart;  
RX MEDLINE=20499358; PubMed=11042143;  
RA Peters T., Dildrop R., Ausmeier K., Ruether U.;  
RT "Organization of mouse Iroquois homeobox genes in two clusters  
RT suggests a conserved regulation and function in vertebrate  
RT development.";  
RL Genome Res. 10:1435-1462(2000).  
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).  
CC -!- SIMILARITY: BELONGS TO THE TALE/IRO HOMEOBOX FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AJ271055; CAC10403.1; -.  
DR HSSP; P40424; 1B72.  
DR MGD; MGI:1927642; Irx6.  
DR InterPro; IPR001064; Crystallin.  
DR InterPro; IPR001356; Homeobox.  
DR InterPro; IPR003893; Iroquois\_homeo.  
DR Pfam; PF00046; homeobox; 1.  
DR ProDom; PD000010; Homeobox; 1.  
DR SMART; SM00389; HOX; 1.  
DR SMART; SM00548; IRO; 1.  
DR PROSITE; PS00027; HOMEOBOX\_1; 1.  
DR PROSITE; PS50071; HOMEOBOX\_2; 1.  
KW Homeobox; DNA-binding; Nuclear protein.  
FT DNA\_BIND 143 205 HOMEOBOX.  
FT DOMAIN 256 267 POLY-GLU.  
SQ SEQUENCE 438 AA; 47242 MW; 4968E1E29CB2A7B2 CRC64;  
  
Query Match 38.7%; Score 39.5; DB 1; Length 438;  
Best Local Similarity 45.0%; Pred. No. 32;  
Matches 9; Conservative 4; Mismatches 4; Indels 3; Gaps 1;  
  
QY 2 TSZPCGAVMILYF---HKWR 18  
I:::| | |||: || |  
Db 325 TAEPCGPTMILHYPSGKPR 344  
  
RESULT 11  
YN32\_ARCFU STANDARD; PRT; 148 AA.  
ID YN32\_ARCFU  
AC O27952;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical protein AF2332.  
GN AF2332.  
OS Archaeoglobus fulgidus.  
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;  
OC Archaeoglobaceae; Archaeoglobus.  
OX NCBI\_Taxid=2234;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;  
RX MEDLINE=98049343; PubMed=9389475;  
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,  
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,  
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,  
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,  
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,  
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,  
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,  
RA Cotton M.D., Spriggs T., Artlich P., Kaine B.P., Sykes S.M.,  
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,  
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,  
RA Venter J.C.;  
RT "The complete genome sequence of the hyperthermophilic, sulphate-  
RT reducing archaeon Archaeoglobus fulgidus.";  
RL Nature 390:364-370(1997).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AE000943; AAB88920.1; -.  
DR TIGR; AF2332; -.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 148 AA; 17384 MW; 8F7FP376CF10669F CRC64;





RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,  
RA Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,  
RA Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,  
RA Joris B., Karamata D., Kasahara Y., Klaer-Bianchard M., Klein C.,  
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,  
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
RA Parro V., Pohl T.M., Portetelle D., Portollik S., Prescott A.M.,  
RA Presecan E., Pujic P., Punelle B., Rapoport G., Rey M., Reynolds S.,  
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,  
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,  
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,  
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.,  
RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*  
RT *subtilis*.";  
RL Nature 390:249-256(1997).  
-1- FUNCTION: PROBABLY PART OF A BINDING-PROTEIN-DEPENDENT TRANSPORT  
SYSTEM YURMO.  
-1- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor  
(Potential).  
-1- SIMILARITY: BELONGS TO THE BACTERIAL EXTRACELLULAR SOLUTE-BINDING  
PROTEIN FAMILY 1.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL: 299120; CAB15250.1; -.  
DR Subtilist; BG14001; YURO.  
DR InterPro: IPR000567; SBP\_bac\_1.  
DR Pfam: PF01547; SBP\_bacterial\_1; 1.  
DR PROSITE: PS00013; PROKAR\_LIPOPROTEIN; 1.  
DR PROSITE: PS01037; SBP\_BACTERIAL\_1; FALSE-NEG.  
KW Hypothetical protein; Transport; Membrane; Lipoprotein; Signal;  
KW Complete proteome.  
FT SIGNAL 1 20 POTENTIAL.  
FT CHAIN 21 422 HYPOTHETICAL ABC TRANSPORTER  
FT LIPID 21 21 EXTRACELLULAR BINDING PROTEIN YURO.  
FT SEQUENCE 422 AA; 48026 MW; 49D67BDC787CF42E CRC64;  
SQ  
very Match 38.2%; Score 39; DB 1; Length 422;  
est Local Similarity 37.5%; Pred. No. 38;  
Matches 6; Conservative 4; Mismatches 6; Indels 0; Gaps 0;  
QY 2 TSZPCGAVMILYFHKW 17  
: 1 1 : 11:1  
Db 25 SSSADGKVTLLKFFHRW 40  
RESULT 15  
KIN2\_YEAST STANDARD; PRT; 1147 AA.  
ID KIN2\_YEAST  
AC P13186; Q12384;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Protein kinase KIN2 (EC 2.7.1.-).  
GN KIN2 OR YLR096W OR L8004.3.  
OS Saccharomycetes cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

OX NCBI\_TaxID=4932;  
RN [1]  
SEQUENCE FROM N.A.  
RX MEDLINE-87317589; PubMed-2957690;  
RA Levin D.E., Hammond C.I., Ralston R.O., Bishop J.M.;  
RT "Two yeast genes that encode unusual protein kinases.";  
RL Proc. Natl. Acad. Sci. U.S.A. 84:6035-6039(1987).  
RN [2]  
SEQUENCE FROM N.A.  
RP Benes V., Rechmann S., Nentwich U., Schwager C., Ansoerge W., Voss H.;  
RA Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
RL [3]  
SEQUENCE FROM N.A.  
RC STRAIN-S288C / AB972;  
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,  
RA Favello A., Fulton L., Gattung S., Greco T., Kirsten J., Kucaba T.,  
RA Hallsworth K., Hawkins J., Hillier L., Jier M., Johnson D.,  
RA Johnston L., Langston Y., Latreille P., Le T., Mardis E., Menezes S.,  
RA Miller N., Nhan M., Pauley A., Peluso D., Rifken L., Riles L.,  
RA Taich A., Trevasakis E., Vignati D., Wilcox L., Wohlman P., Vaudin M.,  
RA Wilson R., Waterston R.;  
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: THIS PROTEIN IS PROBABLY A SERINE/THREONINE PROTEIN  
CC KINASE.  
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
CC NIM1 SUBFAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL: M69018; AAA34723.1; -.  
DR EMBL: 273268; CAA97659.1; -.  
DR EMBL: 273269; CAA97661.1; -.  
DR EMBL: U53876; AAB67540.1; -.  
DR PIR: S42439; S42439.  
DR HSSP: Q63450; 1A06.  
DR SGD: S0004086; KIN2.  
DR InterPro: IPR000719; Euk\_pkinase.  
DR InterPro: IPR001772; Kinase\_Cterm.  
DR InterPro: IPR002290; Ser\_thr\_pkinase.  
DR Pfam: PF00069; pkinase; 1.  
DR Pfam: PF02149; KAI; 1.  
DR ProDom: PD000001; Euk\_pkinase; 1.  
DR SMART: SM00220; S\_TKC; 1.  
DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE: PS00108; PROTEIN\_KINASE\_ST; 1.  
DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
KW Transferase; Serine/threonine-protein kinase; ATP-binding.  
FT DOMAIN 99 377 PROTEIN KINASE.  
FT NP\_BIND 105 113 ATP (BY SIMILARITY).  
FT BINDING 128 128 ATP (BY SIMILARITY).  
FT ACT\_SITE 248 248 BY SIMILARITY.  
FT DOMAIN 527 536 POLY-GLN.  
FT CONFLICT 216 217 OH -> HD (IN REF. 1).  
FT CONFLICT 675 707 QEPLEPEPTYSKSNESIKVPSKSHRTISD ->  
SGTYSSENLOHICQNMKFPKRAIVLYQT (IN  
REF. 1).  
FT CONFLICT 756 758 NAE -> KRO (IN REF. 1).  
FT CONFLICT 805 805 P -> PLSPV (IN REF. 1).  
FT CONFLICT 1034 1037 ATNT -> TTNSI (IN REF. 1).  
FT CONFLICT 1041 1042 NS -> KT (IN REF. 1).  
SQ SEQUENCE 1147 AA; 128338 MW; AC2660BF3CA69600 CRC64;  
Query Match 38.2%; Score 39; DB 1; Length 1147;  
est Local Similarity 43.8%; Pred. No. 1e+02;  
Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;  
QY 1 BTSZPCGAVMILYFHK 16

Db 78 DASRPNCAVELRQFHR 93

Search completed: November 15, 2002, 11:43:30  
Job time : 12 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 15, 2002, 11:43:14 ; Search time 29 Seconds  
(without alignments)  
127.891 Million cell updates/sec

Title: US-09-881-569A-2  
Perfect score: 102  
Sequence: 1 BTRSPCGAVMILYFHKWR 18

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_21:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvivirus:\*  
16: sp\_bacteriophage:\*  
17: sp\_archaeophages:

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID     | Description        |
|------------|-------|-------------|--------|-----------|--------------------|
| 1          | 48    | 47.1        | 155    | 3 Q96U33  | Q96U33 neurospora  |
| 2          | 46    | 45.1        | 326    | 5 Q95XA9  | Q95XA9 caenorhabdi |
| 3          | 45    | 44.1        | 400    | 17 Q8TXF2 | Q8TXF2 methanopyru |
| 4          | 44    | 43.1        | 207    | 1 Q9V229  | Q9V229 uncultured  |
| 5          | 44    | 43.1        | 222    | 17 Q28805 | Q28805 archaeoglob |
| 6          | 44    | 43.1        | 333    | 16 Q8RG17 | Q8RG17 fusobacteri |
| 7          | 44    | 43.1        | 720    | 3 Q8X231  | Q8X231 magnaporthe |
| 8          | 44    | 43.1        | 758    | 6 Q9GL59  | Q9GL59 ovis aries  |
| 9          | 44    | 43.1        | 894    | 10 Q8S4P5 | Q8S4P5 zea mays (m |
| 10         | 43    | 42.2        | 64     | 16 Q8YQ00 | Q8YQ00 anabaena sp |
| 11         | 43    | 42.2        | 992    | 2 Q8RTY6  | Q8RTY6 shewanella  |
| 12         | 43    | 42.2        | 992    | 2 Q8RTY5  | Q8RTY5 shewanella  |
| 13         | 43    | 42.2        | 1239   | 16 Q83868 | Q83868 treponema p |
| 14         | 42    | 41.2        | 79     | 2 Q9F569  | Q9F569 escherichia |
| 15         | 42    | 41.2        | 221    | 16 Q9JS29 | Q9JS29 chlamydia p |
| 16         | 42    | 41.2        | 221    | 16 Q9Z728 | Q9Z728 chlamydia p |

|    |      |      |      |           |                     |
|----|------|------|------|-----------|---------------------|
| 17 | 42   | 41.2 | 298  | 16 Q8ZKY5 | Q8ZKY5 salmonella   |
| 18 | 42   | 41.2 | 302  | 16 Q8Z2N3 | Q8Z2N3 salmonella   |
| 19 | 42   | 41.2 | 443  | 16 Q91667 | Q91667 pseudomonas  |
| 20 | 42   | 41.2 | 602  | 3 Q8TFD3  | Q8TFD3 mycosphaere  |
| 21 | 41.5 | 40.7 | 227  | 2 Q53090  | Q53090 rhodobacter  |
| 22 | 41.5 | 40.7 | 229  | 8 Q63548  | Q63548 drosophila   |
| 23 | 41   | 40.2 | 259  | 5 Q45048  | Q45048 anopheles g  |
| 24 | 41   | 40.2 | 317  | 17 Q97CM0 | Q97CM0 thermoplasma |
| 25 | 41   | 40.2 | 320  | 10 Q945T6 | Q945T6 actinidia d  |
| 26 | 41   | 40.2 | 353  | 10 Q9SK02 | Q9SK02 arabidopsis  |
| 27 | 41   | 40.2 | 505  | 2 Q9R9S3  | Q9R9S3 aeromonas p  |
| 28 | 41   | 40.2 | 558  | 12 Q9YMQ7 | Q9YMQ7 lymantria d  |
| 29 | 41   | 40.2 | 575  | 16 Q927A1 | Q927A1 listeria in  |
| 30 | 41   | 40.2 | 575  | 16 Q8Y3T6 | Q8Y3T6 listeria mo  |
| 31 | 41   | 40.2 | 612  | 16 Q9HU20 | Q9HU20 pseudomonas  |
| 32 | 41   | 40.2 | 616  | 5 Q9N977  | Q9N977 leishmania   |
| 33 | 41   | 40.2 | 746  | 5 Q16992  | Q16992 caenorhabdi  |
| 34 | 40.5 | 39.7 | 497  | 16 Q8XRY6 | Q8XRY6 raistonia s  |
| 35 | 40.5 | 39.7 | 681  | 16 Q8XTJ7 | Q8XTJ7 raistonia s  |
| 36 | 40.5 | 39.7 | 868  | 10 Q9SH67 | Q9SH67 arabidopsis  |
| 37 | 40.5 | 39.7 | 954  | 5 Q95Y23  | Q95Y23 leishmania   |
| 38 | 40.5 | 39.7 | 2133 | 12 Q98203 | Q98203 molluscum c  |
| 39 | 40   | 39.2 | 74   | 2 Q50168  | Q50168 mycobacteri  |
| 40 | 40   | 39.2 | 156  | 5 Q8SUW8  | Q8SUW8 encephalito  |
| 41 | 40   | 39.2 | 199  | 16 Q8U882 | Q8U882 agrobacteri  |
| 42 | 40   | 39.2 | 257  | 10 Q9XFD1 | Q9XFD1 arabidopsis  |
| 43 | 40   | 39.2 | 335  | 2 Q9L610  | Q9L610 thlocapsa r  |
| 44 | 40   | 39.2 | 450  | 2 Q46730  | Q46730 escherichia  |
| 45 | 40   | 39.2 | 476  | 2 Q46731  | Q46731 escherichia  |

ALIGNMENTS

|                       |   |   |        |                                 |
|-----------------------|---|---|--------|---------------------------------|
| RESULT 1              | Q96U33  | PRELIMINARY:                            | PRT:   | 155 AA.                         |
| ID                    | Q96U33  | AC                                      | Q96U33 |                                 |
| DT                    | 01-DEC-2001   | (TREMBLrel. 19, Created)                |        |                                 |
| DT                    | 01-DEC-2001   | (TREMBLrel. 19, Last sequence update)   |        |                                 |
| DT                    | 01-DEC-2001   | (TREMBLrel. 19, Last annotation update) |        |                                 |
| DE                    | Hypothetical 17.5 kda protein.  |   |        |                                 |
| GN                    | B208.030.   |   |        |                                 |
| OS                    | Neurospora crassa.  |   |        |                                 |
| OC                    | Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;        |   |        |                                 |
| OC                    | Sordariales; Sordariaceae; Neurospora.                                |   |        |                                 |
| OX                    | NCBI_TaxID=5141;  |   |        |                                 |
| RN                    | [1]   | SEQUENCE FROM N.A.                      |        |                                 |
| RP                    | Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R., |   |        |                                 |
| RA                    | Nyakatura G., Mewes H.W., Mannhaupt G.;                               |   |        |                                 |
| RL                    | Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.               |   |        |                                 |
| RN                    | [2]   | SEQUENCE FROM N.A.                      |        |                                 |
| RP                    | German Neurospora genome project;                                     |   |        |                                 |
| RA                    | Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.               |   |        |                                 |
| DR                    | EMBL; AL355930; CAD01121.1;   |   |        |                                 |
| KW                    | Hypothetical protein.   |   |        |                                 |
| SQ                    | SEQUENCE 155 AA; 17483 MW; 80E455D3BF0ECEAD CRC64;                    |   |        |                                 |
| Query Match           | 47.1%;  | Score 48;                               | DB 3;  | Length 155;                     |
| Best Local Similarity | 40.0%;  | Pred. No. 1.8;                          |        |                                 |
| Matches               | 6;  | Conservative                            | 4;     | Mismatches 5; Indels 0; Gaps 0; |
| QY                    | 4 ZPCGAVMILYFHKWR 18  |   |        |                                 |
| DB                    | 50 EPCGAVVFPLRNWK 64  |   |        |                                 |
| RESULT 2              | Q95XA9  | PRELIMINARY:                            | PRT:   | 326 AA.                         |
| ID                    | Q95XA9  | AC                                      | Q95XA9 |                                 |

DT 01-DEC-2001 (Tremblrel. 19, Created)  
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE Hypothetical 37.6 kDa protein.  
 GN y20F4.3.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA None;  
 RT "Genome sequence of the nematode C. elegans: a platform for  
 investigating biology. The C. elegans Sequencing Consortium.";  
 RL Science 282:2012-2018(1998).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 RA Waterston R.;  
 RT "Direct Submision.";  
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC093703; AAL00863.1; -.  
 DR InterPro; IPR000980; SH2.  
 DR InterPro; IPR001452; SH3.  
 DR Pfam; PF00017; SH2; 1.  
 DR Pfam; PF00018; SH3; 1.  
 DR PROSITE; PSS0001; SH2; 1.  
 DR PROSITE; PSS0002; SH3; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 326 AA; 37576 MW; 5885F810848586CE CRC64;  
 QY 5 PCGAVMILYFHKW 17  
 Db 61 PCSPSSISHFHKW 73

Query Match 45.1%; Score 46; DB 5; Length 326;  
 Best Local Similarity 53.8%; Pred. No. 7.8;  
 Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

RESULT 3  
 DT 08TXF2 PRELIMINARY; PRT; 400 AA.  
 DT 08TXF2;  
 DT 01-JUN-2002 (Tremblrel. 21, Created)  
 DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE CO dehydrogenase/acetyl-CoA synthase delta subunit (corrinoid Fe-S  
 DE protein).  
 GN CDHD OR MK0722.  
 OS Methanopyrus kandleri.  
 OC Archaea; Euryarchaeota; Methanopyrri; Methanopyrales; Methanopyraceae;  
 OC Methanopyrus.  
 GN NCBI\_TaxID=2320;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AV19 / DSM 6324 / JCM 9639;  
 RX MEDLINE=21927647; PubMed=11930014;  
 RA Slesarev A.I., Mezheva K.V., Makarova K.S., Polushin N.N.,  
 RA Shcherbina O.V., Shakhova V.V., Belova G.I., Aravind L.,  
 RA Natale D.A., Rogozin I.B., Tatusov R.L., Wolf Y.I., Stetter K.O.,  
 RA Mal'kh A.G., Koonin E.V., Kozlovskiy S.A.;  
 RT "The complete genome of hyperthermophile Methanopyrus kandleri AV19  
 RT and monophyly of archaeal methanogens.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649(2002).  
 DR EMBL; AEO10365; AAM01936.1; -.  
 KW Complete proteome.  
 SQ SEQUENCE 400 AA; 44542 MW; 83ADA136E30C07CB CRC64;  
 QY 44.1%; Score 45; DB 17; Length 400;  
 Best Local Similarity 50.0%; Pred. No. 14;

Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
 QY 6 CGAVMILYFHKW 17  
 Db 361 CGADLMMFHPW 372

RESULT 4  
 DT 09V229 PRELIMINARY; PRT; 207 AA.  
 AC 09V229;  
 DT 01-MAY-2000 (Tremblrel. 13, Created)  
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
 DT 01-OCT-2001 (Tremblrel. 18, Last annotation update)  
 DE MCR4 (Fragment).  
 GN MCR4.  
 OS uncultured methanogen ODP8-ME2.  
 OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;  
 OC environmental samples.  
 OX NCBI\_TaxID=95922;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99365980; PubMed=10436927;  
 RA Bidle K.A., Kestner M., Bartlett D.H.;  
 RT "A phylogenetic analysis of microbial communities associated with  
 RT methane hydrate containing marine fluids and sediments in the Cascadia  
 RT margin (ODP site 892B)."  
 RL FEMS Microbiol. Lett. 177:101-108(1999).  
 DR EMBL; AF121100; AAD45632.1; -.  
 DR HSSP; P11558; 1MRO.  
 DR InterPro; IPR003183; MCR\_alpha.  
 DR Pfam; PF02249; MCR\_alpha; 1.  
 DR Pfam; PF02745; MCR\_alpha\_N; 1.  
 FT NON\_TER 1  
 FT NON\_TER 207  
 SQ SEQUENCE 207 AA; 22546 MW; 4A507ECA265D8621 CRC64;  
 QY 43.1%; Score 44; DB 1; Length 207;  
 Best Local Similarity 50.0%; Pred. No. 11;  
 Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

RESULT 5  
 DT 028805 PRELIMINARY; PRT; 222 AA.  
 AC 028805;  
 DT 01-JAN-1998 (Tremblrel. 05, Created)  
 DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE Signal-transducing histidine kinase.  
 GN AFI1467.  
 OS Archaeoglobus fulgidus.  
 OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;  
 OC Archaeoglobaceae; Archaeoglobus.  
 OX NCBI\_TaxID=2234;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;  
 RX MEDLINE=98049343; PubMed=9389475;  
 RA Kleink H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,  
 RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,  
 RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyriides N.C.,  
 RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,  
 RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,  
 RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodak A., Zhou L.,  
 RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,  
 RA Cotton M.D., Spriggs T., Artach P., Kaine B.P., Sykes S.M.,  
 RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,  
 RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,

```
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RL reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).
DR EMBL; AE001001; AAB89779.1; -.
DR TIGR; AF1467; -.
DR InterPro; IPR003594; ATPbind_ATPase.
DR InterPro; IPR003661; His_kinA.
DR InterPro; IPR004359; His_KIN_sig.
DR Pfam; PF02518; HATPase_c; 1.
DR SMART; SM00387; HATPase_c; 1.
DR SMART; SM00388; HSKA; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 222 AA; 25036 MW; 1B7BAE1B9FB70705 CRC64;

Query Match 43.1%; Score 44; DB 17; Length 222;
Best Local Similarity 50.0%; Pred. No. 12;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 3 SZPCGAVMILYF 14
Db 204 NEPCGAVFVMEF 215

RESULT 6
AC Q8RG17; PRELIMINARY; PRT; 333 AA.
DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Iron(III) dicitrate transport system permease protein fecD.
GN FN0306.
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteria; Fusobacterium.
OX NCBI_TaxID=76856;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 25586;
RX MEDLINE-21886394; PubMed-11889109;
RA Kapatal V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
RA Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
RA Vasileva O., Chu L., Kogan Y., Chaga O., Goltsman E., Bernal A.,
RA Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,
RA Fonstein M., Kyripides N., Overbeek R.;
RT "Genome sequence and analysis of the oral bacterium Fusobacterium
RT nucleatum strain ATCC 25586.";
RL J. Bacteriol. 184:2005-2018(2002).
DR EMBL; AE010543; AAL94512.1; -.
KW Complete proteome.
SQ SEQUENCE 333 AA; 36474 MW; FDAFAD8D93A94EB1 CRC64;

Query Match 43.1%; Score 44; DB 16; Length 333;
Best Local Similarity 70.0%; Pred. No. 17;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 9 VMILYFHKWR 18
Db 208 IMILYFLRWR 217

RESULT 7
AC Q8X231; PRELIMINARY; PRT; 720 AA.
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Putative transposase.
OS Magnaporthe grisea (Rice blast fungus) (Pyricularia grisea).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetes Incertae sedis; Magnaporthaceae; Magnaporthe.
OX NCBI_TaxID=148305;
```

```
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-INA168; TRANSPOSON-OCCAN;
RA Kito H., Takahashi Y., Sato J., Sone T., Tomita F.;
RT "Occan, a DNA-type transposon sequence identified from M. grisea.";
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB074754; BAB78449.1; -.
SQ SEQUENCE 720 AA; 81657 MW; 1997D3A9A98E4BE8 CRC64;

Query Match 43.1%; Score 44; DB 3; Length 720;
Best Local Similarity 42.9%; Pred. No. 35;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 5 PCGAVMILYFHKWR 18
Db 107 PCGPVQMNWYGRWR 120

RESULT 8
AC Q9GL59; PRELIMINARY; PRT; 758 AA.
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Vitamin K-dependent gamma-carboxylase.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-LIVER;
RA McCue J.M., Gordy P.W., Cantlon J.D., Baker D.C., Bowen R.A.;
RT "The Sequence of the Ovine Gamma-Carboxylase cDNA.";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF312035; AAG30935.1; -.
DR InterPro; IPR001870; Gamma_carboxylase.
DR InterPro; IPR000504; RNA_rec_mot.
DR PROSITE; PS00030; RRM_RNP_1; UNKNOWN_1.
SQ SEQUENCE 758 AA; 87581 MW; 87F8B58C187B9091 CRC64;

Query Match 43.1%; Score 44; DB 6; Length 758;
Best Local Similarity 46.7%; Pred. No. 37;
Matches 7; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 BTSZPCGAVMILYFH 15
Db 273 DASRPIGFVFSYFH 287

RESULT 9
AC Q8S4P5; PRELIMINARY; PRT; 894 AA.
DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Enhancer of zeste-like protein 2.
GN MEZ2.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Panicoideae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-21948211; PubMed-11950982;
RA Springer N.M., Danilevskaya O.N., Hermon P., Helentjaris T.G.,
RA Phillips R.L., Kaeppler H.F., Kaeppler S.M.;
RT "Sequence Relationships, Conserved Domains, and Expression Patterns
RT for Maize Homologs of the Polycomb Group Genes E(z), esc, and E(Pc).";
```

RL Plant Physiol. 128:1332-1345(2002).  
DR EMBL: AF443597; AAM13421.1; -.  
SO SEQUENCE 894 AA; 99979 MW; 9C2B04E6F8DBA113 CRC64;

Query Match 43.1%; Score 44; DB 10; Length 894;  
Best Local Similarity 38.9%; Pred. No. 43;  
Matches 7; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

OY 1 BTSPCGAVMILYFHKWR 18  
ID :|||:|:|:  
Db 317 ENKKPCGHLCLYRFPQWR 334

RESULT 10

O8Y000 PRELIMINARY; PRT; 64 AA.

DT 01-MAR-2002 (TREMBlrel. 20, Created)  
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
DE Hypothetical protein Asr3770.  
GN ASR3770.  
OS Anabaena sp. (strain PCC 7120).  
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.  
OX NCBI\_TaxID=103690;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21595285; PubMed=11759840;  
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,  
RA Watanabe A., Iriuch M., Ishikawa A., Kawashima K., Kimura T.,  
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,  
RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,  
RA Yasuda M., Tabata S.;  
RT "Complete genomic sequence of the filamentous nitrogen-fixing  
RT cyanobacterium Anabaena sp. strain PCC 7120.";  
RL DNA Res. 8:205-213(2001).  
DR EMBL: AP003594; BAB75469.1; -.  
KW Hypothetical protein; Complete proteome.  
SO SEQUENCE 64 AA; 7403 MW; 92AAD7A9C0F8EF8 CRC64;

Query Match 42.2%; Score 43; DB 16; Length 64;  
Best Local Similarity 58.3%; Pred. No. 5.7;  
Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 6 CGAVMILYFHKW 17  
ID :|||:|:|:  
Db 12 CGGVREKXFHW 23

RESULT 11

O8RTY6 PRELIMINARY; PRT; 992 AA.

DT 01-JUN-2002 (TREMBlrel. 21, Created)  
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE Sphingolipid ceramide N-deacylase.  
OS Shewanella algae.  
OC Bacteria; Proteobacteria; gamma subdivision; Alteromonadaceae;  
OC Shewanella.  
OX NCBI\_TaxID=38313;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=G8;  
RA Furusato M., Sueyoshi N., Mitsutake S., Sakaguchi K., Kita K.,  
RA Okino N., Ichinose S., Omori A., Ito M.;  
RT "Molecular Cloning and Characterization of Sphingolipid Ceramide N-  
RT deacylase from a Marine Bacterium, Shewanella alga G8.";  
RL J. Biol. Chem. 0:0-0(2002).  
DR EMBL: AB079849; BAB85110.1; -.  
SO SEQUENCE 992 AA; 109848 MW; CECBC6CC7A2C6953 CRC64;

Query Match 42.2%; Score 43; DB 2; Length 992;

Best Local Similarity 34.8%; Pred. No. 70;  
Matches 8; Conservative 4; Mismatches 5; Indels 6; Gaps 1;

OY 2 TSZPCGAVMILYFH-----KWR 18  
ID :|||:|:|:  
Db 967 SGEQAGALILYHHGDSNQWK 989

RESULT 12

O8RTY5 PRELIMINARY; PRT; 992 AA.

DT 01-JUN-2002 (TREMBlrel. 21, Created)  
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE Sphingolipid ceramide N-deacylase.  
OS Shewanella algae.  
OC Bacteria; Proteobacteria; gamma subdivision; Alteromonadaceae;  
OC Shewanella.  
OX NCBI\_TaxID=38313;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NS589;  
RA Sueyoshi N., Furusato M., Okino N., Mitsutake S., Ito M.;  
RT "Molecular cloning and characterization of sphingolipid ceramide N-  
RT deacylase from a marine bacterium, Shewanella alga NS589.";  
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AB079850; BAB85111.1; -.  
SO SEQUENCE 992 AA; 109862 MW; B7611F6BE9DAE8BA CRC64;

Query Match 42.2%; Score 43; DB 2; Length 992;  
Best Local Similarity 34.8%; Pred. No. 70;  
Matches 8; Conservative 4; Mismatches 5; Indels 6; Gaps 1;

OY 2 TSZPCGAVMILYFH-----KWR 18  
ID :|||:|:|:  
Db 967 SGEQAGALILYHHGDSNQWK 989

RESULT 13

O83868 PRELIMINARY; PRT; 1239 AA.

DT 01-NOV-1998 (TREMBlrel. 08, Created)  
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE ATP-dependent nuclease, subunit A, putative.  
GN TP0898.  
OS Treponema pallidum.  
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.

OX NCBI\_TaxID=160;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NICHOLS;  
RX MEDLINE=98332770; PubMed=9665876;  
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,  
RA Dodson R., Gwin M., Hickey E.K., Clayton R., Ketchum K.A.,  
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,  
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,  
RA McDonald L., Artach P., Bowman C., Cotton M.D., Fujii C., Garland S.,  
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,  
RA Venter J.C.;  
RT "Complete genome sequence of Treponema pallidum, the syphilis  
RT spirochete.";  
RL Science 281:375-388(1998).  
DR EMBL: AE001259; AAC65853.1; -.  
DR HSSP: P09980; 1UAA.  
DR TIGR: TP0898; -.  
DR InterPro: IPR000212; UVRD-helicase.  
DR Pfam: PF00580; UVRD-helicase; 1.  
KW Complete proteome.  
SO SEQUENCE 1239 AA; 137095 MW; DCCDA3B1F267597 CRC64;

Query Match 42.2%; Score 43; DB 16; Length 1239;  
Best Local Similarity 53.3%; Pred. No. 86;  
Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy 2 TSZPCGAVMILYFHK 16  
Db 833 TDRACGAVOIMSVHK 847

## RESULT 14

09F569 PRELIMINARY; PRT; 79 AA.  
AC 09F569;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE YajB protein.  
GN YajB.  
OS Escherichia coli.  
OG Plasmid R721.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_Taxid-562;  
[1]

SEQUENCE FROM N.A.  
STRAIN-K-12; TRANSPOSON-TN7;  
RA Sampa G., Motomura K., Masuda S., Yamaguchi T., Ando K., Olshi T.,  
RA Furuya N., Komano T., Mizobuchi K.;  
RT "Organization and diversification of plasmid genomes: complete  
RL nucleotide sequence of the R721 genome.";  
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
[2]

SEQUENCE FROM N.A.  
STRAIN-K-12; TRANSPOSON-TN7;  
RC MEDLINE-93015772; PubMed-1400257;  
RX Kim S., Komano T.;  
RA "Nucleotide sequence of the R721 shufflon.";  
RT J. Bacteriol. 174:7053-7058(1992).  
DR EMBL; AP002527; BAB12599.1;  
KW Plasmid.  
SQ SEQUENCE 79 AA; 8923 MW; 091817FE76671D65 CRC64;

Query Match 41.2%; Score 42; DB 2; Length 79;  
Best Local Similarity 33.3%; Pred. No. 10;  
Matches 8; Conservative 4; Mismatches 4; Indels 8; Gaps 1;

Qy 3 SZPCGA-----VMILYFHKWR 18  
Db 47 SMPCRALHECNCLYLMVFFFEWR 70

## RESULT 15

09JS29

PRELIMINARY;

PRT; 221 AA.

AC 09JS29;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)  
DE SET domain protein.  
GN CPJ0878 OR CP0991.  
OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
OX NCBI\_Taxid-83558;  
[1]

SEQUENCE FROM N.A.

STRAIN-AR39;

MEDLINE-20150255; PubMed-10684935;

RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,  
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,  
RA Gwin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,  
RA Eisen J., Fraser C.M.;  
RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia

RT pneumoniae AR39.";  
RL Nucleic Acids Res. 28:1397-1406(2000).  
RN [2]

SEQUENCE FROM N.A.

RC STRAIN-J138;  
RX MEDLINE-20330349; PubMed-10871362;  
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,  
RA Shiba T., Ishi K., Hattori M., Kuhara S., Nakazawa T.;  
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138  
RT from Japan and CML029 from USA.";  
RL Nucleic Acids Res. 28:2311-2314(2000).  
DR EMBL; AE002257; AAF38770.1;  
DR EMBL; AP002548; BAA99086.1;  
DR TIGR; CP0991;  
DR InterPro; IPR001214; SET.  
DR Pfam; PF00856; SET; 1.  
DR SMART; SM00317; SET; 1.  
DR PROSITE; PS50280; SET; 1.  
SQ SEQUENCE 221 AA; 25889 MW; 4F6750E73FA08F6B CRC64;

Query Match 41.2%; Score 42; DB 16; Length 221;  
Best Local Similarity 35.3%; Pred. No. 26;  
Matches 6; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

Qy 2 TSZPCGAVMILYFHKWR 18  
Db 5 TTEPCSSIHISLNDWR 21

Search completed: November 15, 2002, 11:45:11  
Job time : 31 secs

